

リーディングDAT養成コース

課題レポート L-B1

田中 涼

2021. 01. 08

課題1

① 対象国を「イギリス」とし、イギリスにおける訪日外客数の月次推移を状態空間モデルで解析します。

基本構造時系列モデルのパターン

- ローカルレベルモデル + 固定季節変動
- ローカルレベルモデル + 可変季節変動
- 平滑化トレンドモデル + 固定季節変動
- 平滑化トレンドモデル + 可変季節変動
- ローカル線形トレンドモデル + 固定季節変動
- ローカル線形トレンドモデル + 可変季節変動
- ローカルレベルモデル + 固定季節変動 (対数変換あり)
- ローカルレベルモデル + 可変季節変動 (対数変換あり)
- 平滑化トレンドモデル + 固定季節変動 (対数変換あり)
- 平滑化トレンドモデル + 可変季節変動 (対数変換あり)
- ローカル線形トレンドモデル + 固定季節変動 (対数変換あり)
- ローカル線形トレンドモデル + 可変季節変動 (対数変換あり)

```
irow <- "英國"
label_list <- c()
# ローカルレベルモデル+固定季節変動
mod1 <- SSModel(jnto[, irow] ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=0), H=NA)
fit1 <- fitSSM(mod1, numeric(2), method = "SANN")
kfs1 <- KFS(fit1$model, smoothing=c("state", "mean", "disturbance"))
logLik1 <- kfs1$logLik - sum(kfs1$Finf>0) * log(2*pi)/2
AIC1 <- -2*logLik1 + 2*( 2 + 12 )
label_list <- "ローカル_固定"
```

```
# ローカルレベルモデル+可変季節変動
mod2 <- SSModel(jnto[, irow] ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=NA), H=NA)
fit2 <- fitSSM(mod2, numeric(3), method = "SANN")
kfs2 <- KFS(fit2$model, smoothing=c("state", "mean", "disturbance"))
logLik2 <- kfs2$logLik - sum(kfs2$Finf > 0) * log(2*pi)/2
AIC2 <- -2*logLik2 + 2*( 3 + 12 )
label_list <- c(label_list, "ローカル_可変")
```

```
# 平滑化トレンドモデル+固定季節変動
mod3 <- SSModel(jnto[, irow] ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=0), H=NA)
fit3 <- fitSSM(mod3, numeric(2), method = "SANN")
kfs3 <- KFS(fit3$model, smoothing=c("state", "mean", "disturbance"))
logLik3 <- kfs3$logLik - sum(kfs3$Finf > 0) * log(2*pi)/2
AIC3 <- -2*logLik3 + 2*( 2 + 13 )
label_list <- c(label_list, "平滑_固定")
```

```
# 平滑化トレンドモデル+可変季節変動
mod4 <- SSModel(jnto[, irow] ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit4 <- fitSSM(mod4, numeric(3), method = "SANN")
kfs4 <- KFS(fit4$model, smoothing=c("state", "mean", "disturbance"))
logLik4 <- kfs4$logLik - sum(kfs4$Finf > 0) * log(2*pi)/2
AIC4 <- -2*logLik4 + 2*( 3 + 13 )
label_list <- c(label_list, "平滑_可変")
```

```
# ローカル線形トレンドモデル+固定季節変動
mod5 <- SSModel(jnto[, irow] ~ SSMtrend(2, Q=list(NA, NA)) + SSMseasonal(12, Q=0), H=NA)
fit5 <- fitSSM(mod5, numeric(3), method = "SANN")
kfs5 <- KFS(fit5$model, smoothing=c("state", "mean", "disturbance"))
logLik5 <- kfs5$logLik - sum(kfs5$Finf > 0) * log(2*pi)/2
AIC5 <- -2*logLik5 + 2*( 3 + 13 )
label_list <- c(label_list, "ローカル線形_固定")
```

```
# ローカル線形トレンドモデル+可変季節変動
mod6 <- SSModel(jnto[, irow] ~ SSMtrend(2, Q=list(NA, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit6 <- fitSSM(mod6, numeric(4), method = "SANN")
kfs6 <- KFS(fit6$model, smoothing=c("state", "mean", "disturbance"))
logLik6 <- kfs6$logLik - sum(kfs6$Finf > 0) * log(2*pi)/2
AIC6 <- -2*logLik6 + 2*( 4 + 13 )
label_list <- c(label_list, "ローカル線形_可変")
```

```
# ローカルレベルモデル+固定季節変動（対数変換あり）
mod7 <- SSModel(log(jnto[, irow]) ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=0 ), H=NA)
fit7 <- fitSSM(mod7, numeric(2), method = "SANN")
kfs7 <- KFS(fit7$model)
logLik7 <- kfs7$logLik - sum(kfs7$Finf>0) * log(2*pi)/2
AIC7 <- -2*logLik7 + 2*( 2 + 12 ) + 2*sum(log(jnto[, irow]))
label_list <- c(label_list, "ローカル_固定_log")
```

```
# mod8 ローカルレベルモデル+可変季節変動（対数変換あり）
mod8 <- SSModel(log(jnto[, irow]) ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=NA), H=NA)
fit8 <- fitSSM(mod8, numeric(3), method = "SANN")
kfs8 <- KFS(fit8$model, smoothing=c("state", "mean", "disturbance"))
logLik8 <- kfs8$logLik - sum(kfs8$Finf > 0) * log(2*pi)/2
AIC8 <- -2*logLik8 + 2*( 3 + 12 ) + 2*sum(log(jnto[, irow]))
label_list <- c(label_list, "ローカル_可変_log")
```

```
# mod9 平滑化トレンドモデル+固定季節変動（対数変換あり）
mod9 <- SSModel(log(jnto[, irow]) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=0 ), H=NA)
fit9 <- fitSSM(mod9, numeric(2), method = "SANN")
kfs9 <- KFS(fit9$model, smoothing=c("state", "mean", "disturbance"))
logLik9 <- kfs9$logLik - sum(kfs9$Finf > 0) * log(2*pi)/2
AIC9 <- -2*logLik9 + 2*( 2 + 13 ) + 2*sum(log(jnto[, irow]))
label_list <- c(label_list, "平滑_固定_log")
```

```
# mod10 平滑化トレンドモデル+可変季節変動（対数変換あり）
mod10 <- SSModel(log(jnto[, irow]) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit10 <- fitSSM(mod10, numeric(3), method = "SANN")
kfs10 <- KFS(fit10$model, smoothing=c("state", "mean", "disturbance"))
logLik10 <- kfs10$logLik - sum(kfs10$Finf > 0) * log(2*pi)/2
AIC10 <- -2*logLik10 + 2*( 3 + 13 ) + 2*sum(log(jnto[, irow]))
label_list <- c(label_list, "平滑_可変_log")
```

```
# mod11 ローカル線形トレンドモデル+固定季節変動（対数変換あり）
mod11 <- SSModel(log(jnto[, irow]) ~ SSMtrend(2, Q=list(NA, NA)) + SSMseasonal(12, Q=0), H=NA)
fit11 <- fitSSM(mod11, numeric(3), method = "SANN")
kfs11 <- KFS(fit11$model, smoothing=c("state", "mean", "disturbance"))
logLik11 <- kfs11$logLik - sum(kfs11$Finf > 0) * log(2*pi)/2
AIC11 <- -2*logLik11 + 2*( 3 + 13 ) + 2*sum(log(jnto[, irow]))
label_list <- c(label_list, "ローカル線形_固定_log")
```

```
# mod12 ローカル線形トレンドモデル+可変季節変動（対数変換あり）
mod12 <- SSModel(log(jnto[, irow]) ~ SSMtrend(2, Q=list(NA, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit12 <- fitSSM(mod12, numeric(4), method = "SANN")
kfs12 <- KFS(fit12$model, smoothing=c("state", "mean", "disturbance"))
logLik12 <- kfs12$logLik - sum(kfs12$Finf > 0) * log(2*pi)/2
AIC12 <- -2*logLik12 + 2*( 4 + 13 ) + 2*sum(log(jnto[, irow]))
label_list <- c(label_list, "ローカル線形_可変_log")
```

- 最も良いモデルとそのAICは以下の通り。

```
AIC_list <- c(AIC1, AIC2, AIC3, AIC4, AIC5, AIC6, AIC7, AIC8, AIC9, AIC10, AIC11, AIC12)
AIC_df <- data.frame(AIC_list, row.names = label_list)
Mod_name <- paste("最良モデル:", rownames(AIC_df)[which.min(AIC_df$AIC_list)])
min_AIC <- paste("AIC:", AIC_df[which.min(AIC_df$AIC_list),])
print(paste(Mod_name, min_AIC))
```

```
## [1] "最良モデル: ローカル線形_可変 , AIC: 3338.56698180584"
```

- なお、それぞれのモデルのAICは下表の通り。

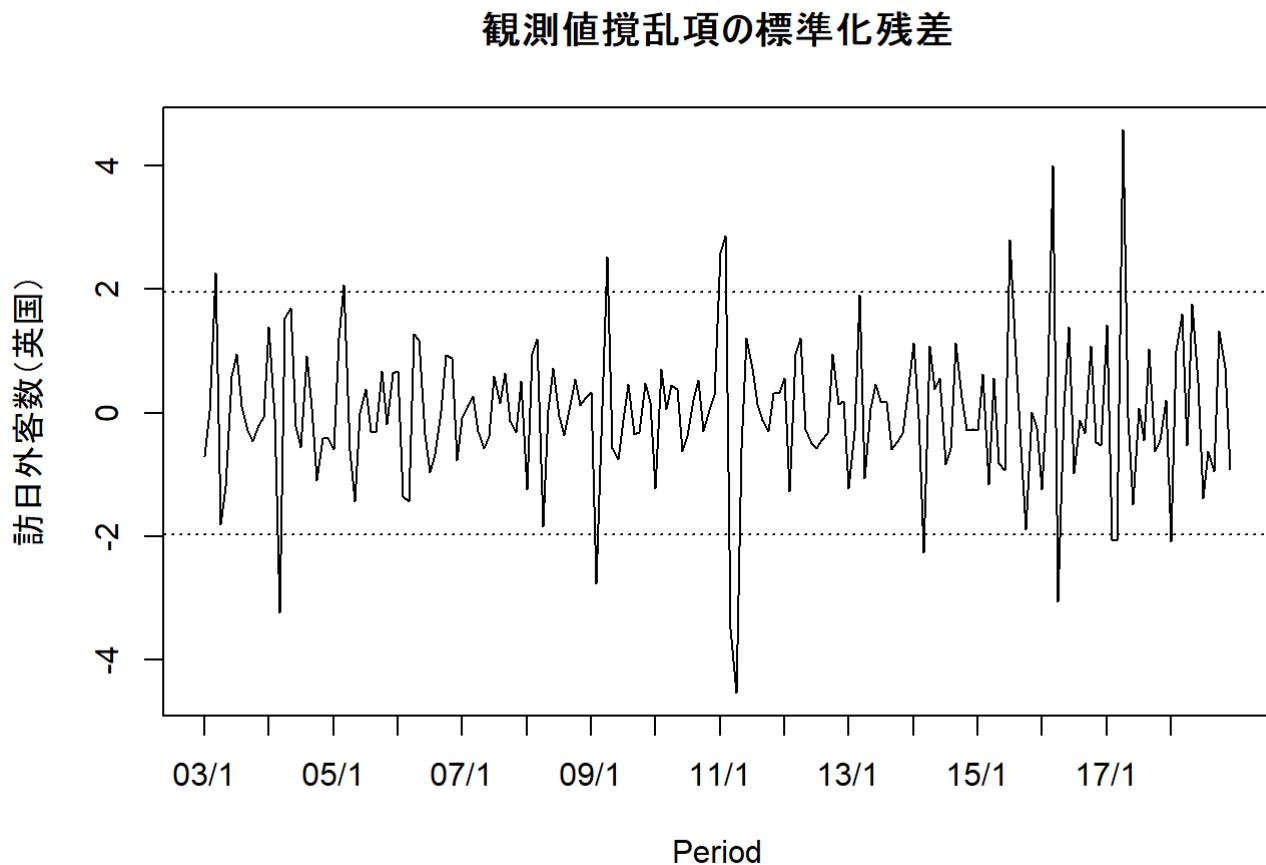
```
print(AIC_df)
```

	AIC_list
## ローカル_固定	3726.663
## ローカル_可変	3363.105
## 平滑_固定	3371.380
## 平滑_可変	3345.256
## ローカル線形_固定	3372.949
## ローカル線形_可変	3338.567
## ローカル_固定_log	3568.722
## ローカル_可変_log	3558.723
## 平滑_固定_log	3603.288
## 平滑_可変_log	3600.480
## ローカル線形_固定_log	3583.510
## ローカル線形_可変_log	3573.410

② 続いて、①で最も良いAICを示したローカルレベルモデル+可変季節変動の残差分析を行う

- 観測値搅乱項の残差プロットはこちら

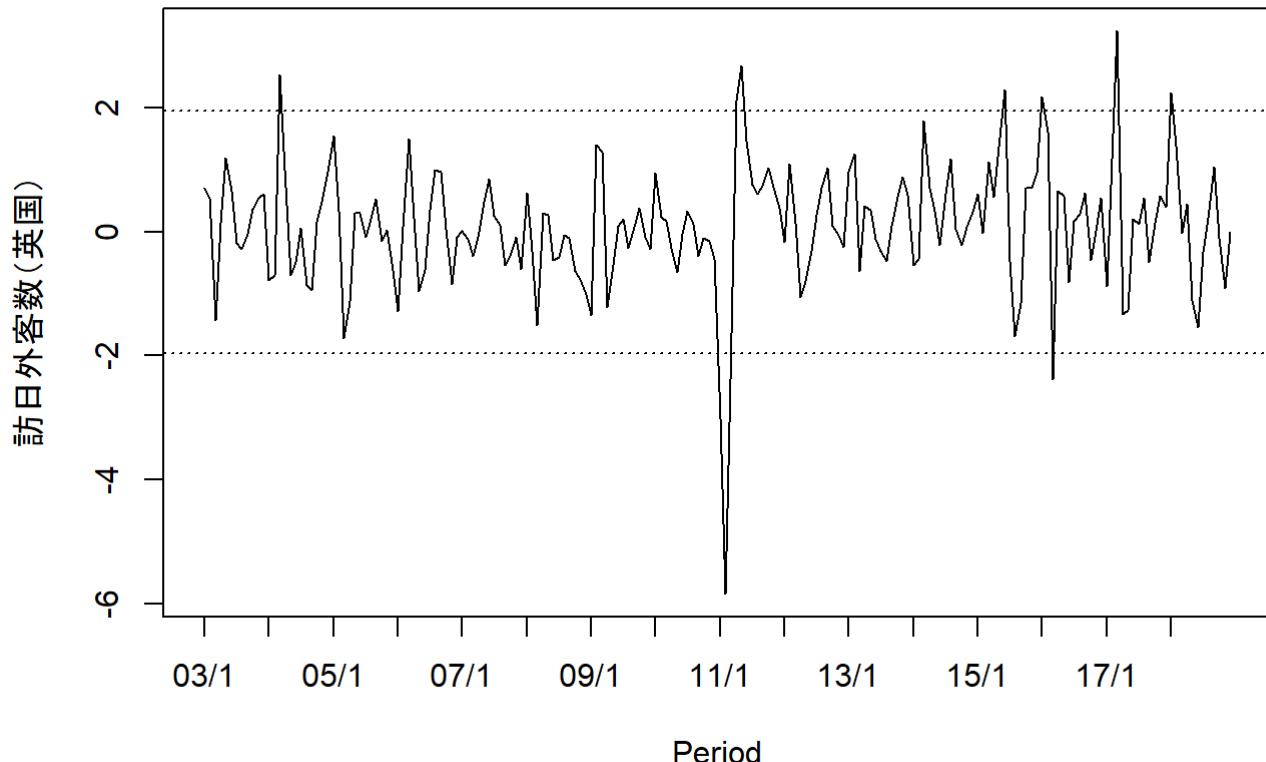
```
plot(rstandard(kfs2, "pearson"), xaxt="n", xlab="Period", ylab="訪日外客数（英国）", main="観測値搅乱項の標準化残差")
abline(h=c(-1.96, 1.96), lty=3)
axis(side=1, at=1+0:15*12, labels=c("03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1", "16/1", "17/1", "18/1"))
```



- 状態搅乱項の残差プロットはこちら

```
plot(rstandard(kfs2, "state") [,1], xaxt="n", xlab="Period", ylab="訪日外客数（英国）", main="状態搅乱項の標準化残差")
abline(h=c(-1.96, 1.96), lty=3)
axis(side=1, at=1+0:15*12, labels=c("03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1", "16/1", "17/1", "18/1"))
```

状態搅乱項の標準化残差



残差分析の結果から分かること

- 観測値搅乱項について、2011年3月付近に負の異常残差がみられる。
 - これは、東日本大震災による一時的な減少と推測する。
 - 打ち手としては以下の2点を考案する。
 - 异常値として除外する
 - 短期変動として切り出す
- 観測値搅乱項において、2015年9月あたりを境に定期的に正の異常残差がみられる。
 - 状態搅乱項の残差の推移など総合的に検討すると、訪日外客数の上昇トレンドが発生していると考えられる。
 - 打ち手としては以下を考案する。
 - 判断基準（※）を用いて期間を設定し、それ以降の期間で上昇トレンドの成分を加える
(干渉変数で構造を変える) ※判断基準について：昨対比、前月比あたりの伸び率を算出し、可視化。違いが現れるようならそれを基準にしたい。

2011年3月付近のデータを異常値として除外したモデルを構築する。

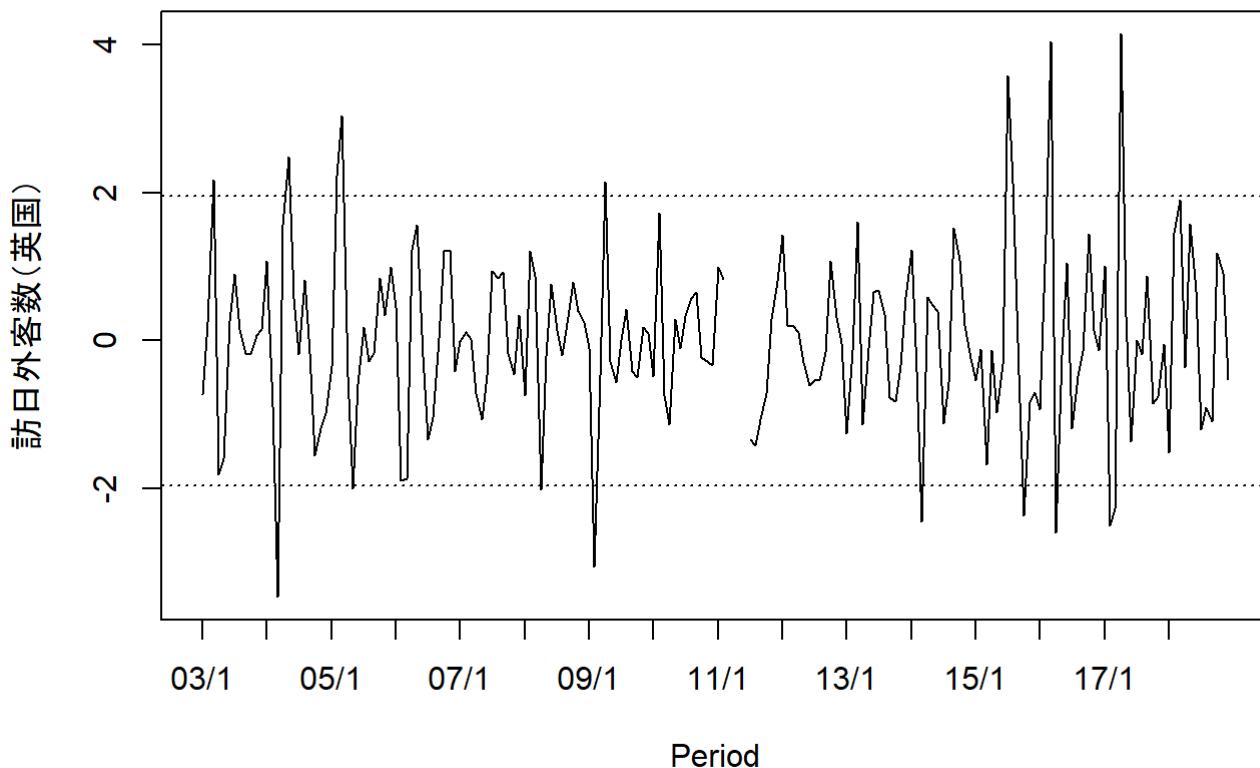
```
# 2011年3~5月のデータを除外
jntoNA <- jnto
jntoNA[jntoNA[, "年月"] %in% c("2011年3月", "2011年4月", "2011年5月", "2011年6月") == TRUE, 2:22]
<- NA

mod_NA2 <- SSModel(jntoNA[, irow] ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=NA), H=NA)
fit_NA2 <- fitSSM(mod_NA2, numeric(3), method = "SANN")
kfs_NA2 <- KFS(fit_NA2$model, smoothing=c("state", "mean", "disturbance"))
logLik_NA2 <- kfs_NA2$logLik - sum(kfs_NA2$Finf > 0) * log(2*pi)/2
AIC_NA2 <- -2*logLik_NA2 + 2*(3 + 12)
```

- 観測値搅乱項の残差プロットはこちら

```
plot(rstandard(kfs_NA2, "pearson"), xaxt="n", xlab="Period", ylab="訪日外客数（英国）", main="観測値搅乱項の標準化残差")
abline(h=c(-1.96, 1.96), lty=3)
axis(side=1, at=1+0:15*12, labels=c("03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1", "16/1", "17/1", "18/1"))
```

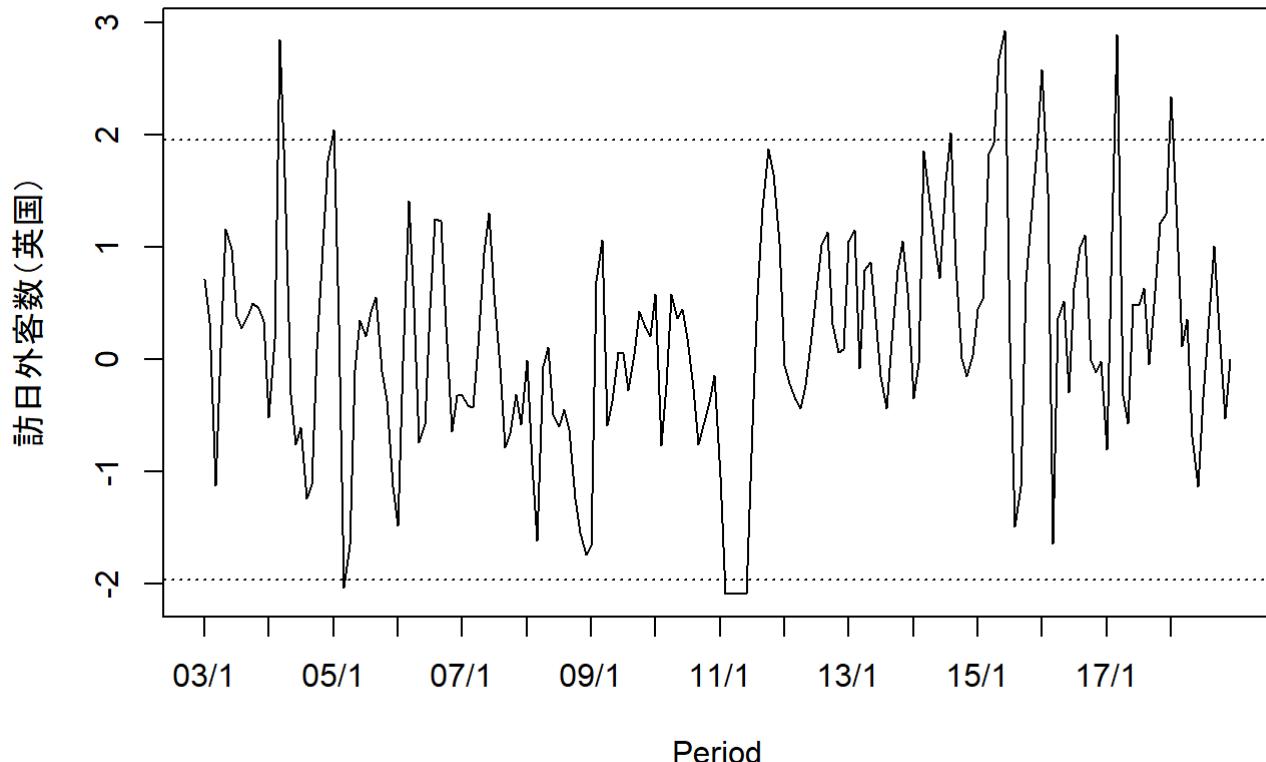
観測値搅乱項の標準化残差



- 状態搅乱項の残差プロットはこちら

```
plot(rstandard(kfs_NA2, "state") [, 1], xaxt="n", xlab="Period", ylab="訪日外客数（英国）", main="状態搅乱項の標準化残差")
abline(h=c(-1.96, 1.96), lty=3)
axis(side=1, at=1+0:15*12, labels=c("03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1", "16/1", "17/1", "18/1"))
```

状態搅乱項の標準化残差



- 2つの残差プロットを見ると、観測値搅乱項の残差において、2010年代前半ごろまでの極端な異常値はなくなったように見える
- しかしながら、依然として2010年代の上昇トレンド傾向が残差の異常値として現れているため、これらに対しても対応が必要なことが伺える。

短期変動成分を加えたモデルを構築する。

- 2011年春ごろの観測値を除外したデータに、短期変動成分を加える。

```
# model2に定常AR成分を導入
mod13_NA <- SSMmodel(log(jntoNA[, irow]) ~ SSMtrend(1, Q= NA) + SSMseasonal(12, Q=NA) + SSMarima
(ar = 0, Q = NA), H = NA)

updatefn <- function(pars, model) {
  SSMmodel(log(jntoNA[, irow]) ~ SSMtrend(1, Q = exp(pars[1]))
    + SSMseasonal(12, Q=exp(pars[2]))
    + SSMarima(ar = artransform(pars[3]), Q = exp(pars[4])),
    H = exp(pars[5]))}

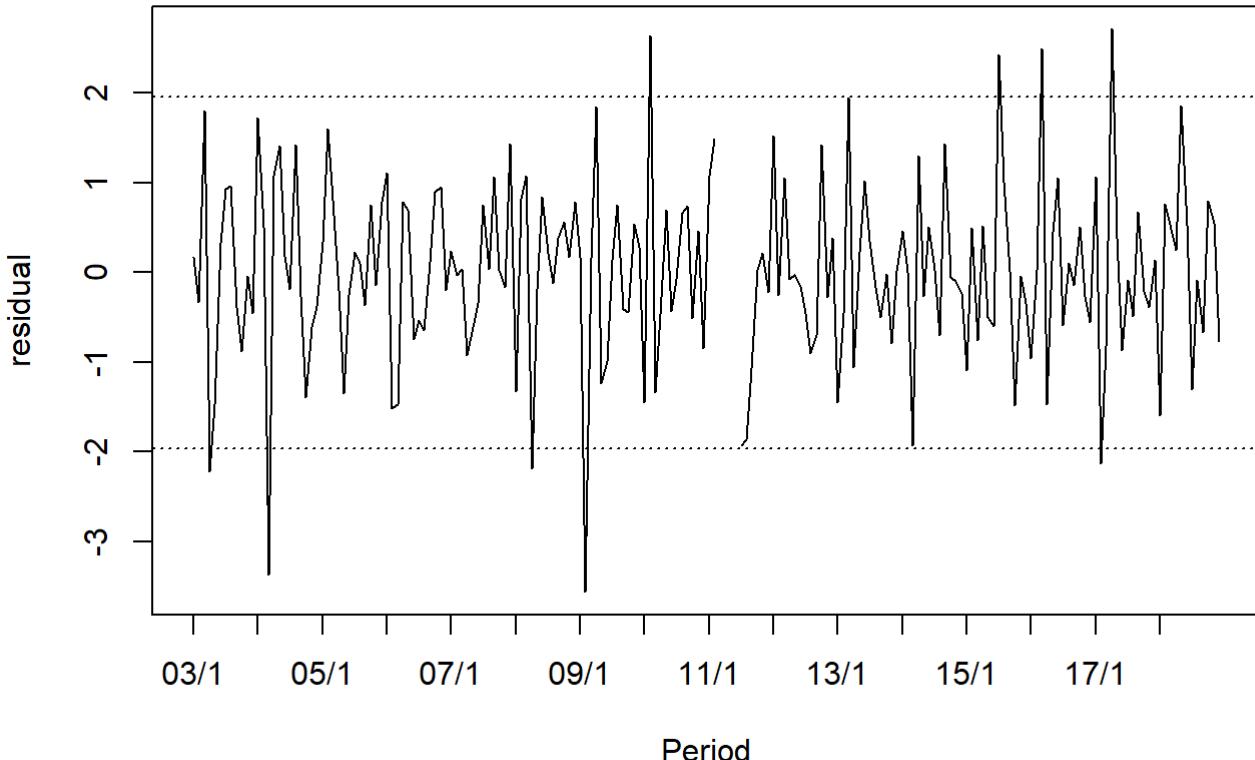
fit13_NA <- fitSSM(mod13_NA, numeric(5), updatefn, method = "BFGS")
kfs13_NA <- KFS(fit13_NA$model, smoothing=c("state", "mean", "disturbance"))
logLik13_NA <- kfs13_NA$logLik - sum(kfs13_NA$Finf > 0) * log(2*pi)/2
AIC13_NA <- -2*logLik13_NA + 2*( 3 + 12 )

# 推定されたAR係数を確認
tail(fit13_NA$model$T, 1)
```

```
## [1] 0.05661548
```

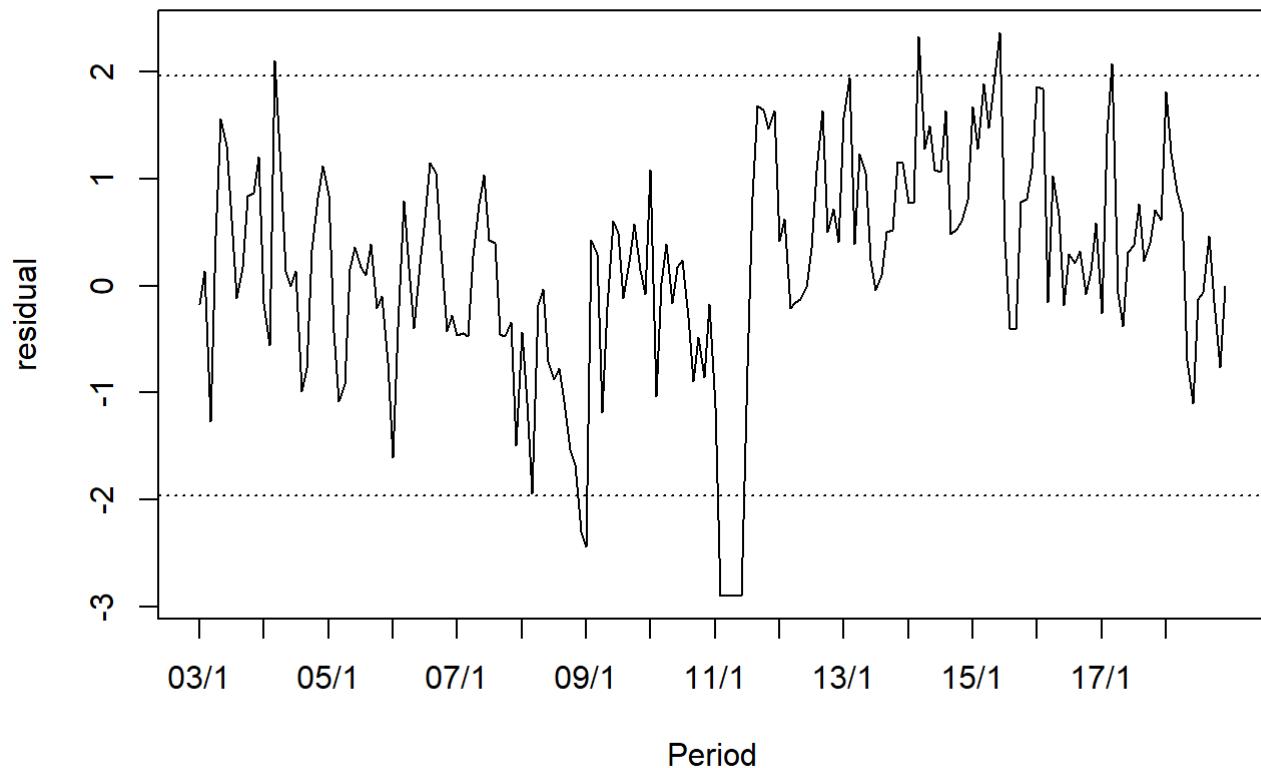
- 観測値搅乱項の残差プロットはこちら

```
plot(rstandard(kfs13_NA, "pearson"), xaxt="n", xlab="Period", ylab="residual")
abline(h=c(-1.96, 1.96), lty=3)
axis(side=1, at=1+0:15*12, labels=c("03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1", "16/1", "17/1", "18/1"))
```



- 状態搅乱項の残差プロットはこちら

```
plot(rstandard(kfs13_NA, "state") [, 1], xaxt="n", xlab="Period", ylab="residual")
abline(h=c(-1.96, 1.96), lty=3)
axis(side=1, at=1+0:15*12, labels=c("03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1", "16/1", "17/1", "18/1"))
```



課題2

準備

- 2011年をNAとするデータセットを用意する。
- 講義同様、対象年齢を20~60歳とする。また、解析時の目的変数は対数変換を施したものを使用する。

解析

- 女性の年齢別死亡率
 - 講義で履修した動的因子モデル（1因子/2因子）を用いて解析する。
- 女性の1因子モデル

```
num_params <- ncol(lmtxf)
# 動的因子モデル（1因子ローカルレベルモデル_女性死亡率対数）
mod_lmtxf1 <- SSModel(lmtxf~SSMtrend(1, Q=1, "common") + SSMtrend(1, Q=0, "distinct"))

updatefn1 <- function(pars, mod_lmtxf1) {
  mod_lmtxf1 <- SSModel(lmtxf~SSMtrend(1, Q=1, "common") + SSMtrend(1, Q=0, "distinct"))
  mod_lmtxf1$Z[, 1, ] <- pars[1:num_params] # length(unique(mtx$Age))
  diag(mod_lmtxf1$H[, , ]) <- exp(pars[num_params+1]) # length(unique(mtx$Age))+1]
  mod_lmtxf1
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxf1 <- fitSSM(mod_lmtxf1, c(rep(0.1, num_params), -5), updatefn1, method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_time - strt_time))
```

```
## [1] "calculation time is:1.03243321577708"
```

```
conf_lmtxf1 <- predict(fit_lmtxf1$model, interval = "confidence", level=0.95)
pre_lmtxf1 <- predict(fit_lmtxf1$model, interval = "prediction", level=0.95)
```

- 1因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 err_lmtxf を算出する。

```
un_2011 <- 2011-1947+1
bet_2011 <- nrow(mtx_under100[mtx_under100$Year=="2011", ])

obs_lmtxf <- lmtxf_nonNA

pred_lmtxf1 <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxf1[nm] <- data.frame(pre_lmtxf1[nm])[un_2011, 1]
}

err_lmtxf <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxf[num_vec] <- pred_lmtxf1[num_vec] - obs_lmtxf[num_vec]
}
```

- 女性の2因子モデル

```
# 動的因子モデル（1因子ローカルレベルモデル_女性死亡率対数）
mod_lmtxf_2fac <- SSModel(lmtxf~SSMtrend(1, Q=1, "common", state_names="factor1_level")
                           +SSMtrend(1, Q=1, "common", state_names="factor2_level")
                           +SSMtrend(1, Q=0, "distinct"))
param2 <- 2*num_params

updatefn4 <- function(pars, mod_lmtxf_2fac) {
  mod_lmtxf_2fac <- SSModel(lmtxf~SSMtrend(1, Q=1, "common", state_names="factor1_level")
                           +SSMtrend(1, Q=1, "common", state_names="factor2_level")
                           +SSMtrend(1, Q=0, "distinct"))

  mod_lmtxf_2fac$Z[, "factor1_level", ] <- pars[1:41]
  mod_lmtxf_2fac$Z[, "factor2_level", ] <- pars[42:82]
  diag(mod_lmtxf_2fac$H[, , ]) <- exp(pars[83])
  mod_lmtxf_2fac
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxf_2fac <- fitSSM(mod_lmtxf_2fac, c(fit_lmtxf1$opt$par[1:41], -20:20/200, -6), updatefn4, method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_timestrt_time))
```

```
## [1] "calculation time is:39.0279831886292"
```

```
conf_lmtxf_2fac <- predict(fit_lmtxf_2fac$model, interval = "confidence", level=0.95)
pre_lmtxf_2fac <- predict(fit_lmtxf_2fac$model, interval = "prediction", level=0.95)
```

- 2因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 err_lmtxf_2fac を算出する。

```
obs_lmtxf_2fac <- lmtxf_nonNA

pred_lmtxf_2fac <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxf_2fac[nm] <- data.frame(pre_lmtxf_2fac[nm]) [un_2011, 1]
}

err_lmtxf_2fac <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxf_2fac[num_vec] <- pred_lmtxf_2fac[num_vec] - obs_lmtxf_2fac[num_vec]
}
```

- 男性の1因子モデル

```
# 動的因子モデル（1因子ローカルレベルモデル_男性死亡率対数）
mod_lmtxm1 <- SSModel(lmtxm~SSMtrend(1, Q=1, "common")+SSMtrend(1, Q=0, "distinct"))

updatefn4 <- function(pars, mod_lmtxm1) {
  mod_lmtxm1 <- SSModel(lmtxm~SSMtrend(1, Q=1, "common")+SSMtrend(1, Q=0, "distinct"))
  mod_lmtxm1$Z[, 1, ] <- pars[1:num_params]#length(unique(mtx$Age))]
  diag(mod_lmtxm1$H[, , ]) <- exp(pars[num_params+1])#length(unique(mtx$Age))+1]
  mod_lmtxm1
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxm1 <- fitSSM(mod_lmtxm1, c(rep(0.1, num_params), -5), updatefn4, method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_timestrt_time))
```

```
## [1] "calculation time is:51.1458721160889"
```

```
conf_mtxm1 <- predict(fit_lmtxm1$model, interval = "confidence", level=0.95)
pre_mtxm1 <- predict(fit_lmtxm1$model, interval = "prediction", level=0.95)
```

- 1因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 err_lmtxm を算出する。

```
obs_lmtxm <- lmtxm_nonNA

pred_lmtxm1 <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxm1[nm] <- data.frame(pre_mtxm1[nm])[un_2011, 1]
}

err_lmtxm <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxm[num_vec] <- pred_lmtxm1[num_vec] - obs_lmtxm[num_vec]
}
```

- 男性の2因子モデル

```
# 動的因子モデル（1因子ローカルレベルモデル_女性死亡率対数）
mod_lmtxm_2fac <- SSModel(lmtxm~SSMtrend(1, Q=1, "common", state_names="factor1_level")
                           +SSMtrend(1, Q=1, "common", state_names="factor2_level")
                           +SSMtrend(1, Q=0, "distinct"))
param2 <- 2*num_params

updatefn5 <- function(pars, mod_lmtxm_2fac) {
  mod_lmtxm_2fac <- SSModel(lmtxm~SSMtrend(1, Q=1, "common", state_names="factor1_level")
                           +SSMtrend(1, Q=1, "common", state_names="factor2_level")
                           +SSMtrend(1, Q=0, "distinct"))
  mod_lmtxm_2fac$Z[, "factor1_level", ] <- pars[1:41]
  mod_lmtxm_2fac$Z[, "factor2_level", ] <- pars[42:82]
  diag(mod_lmtxm_2fac$H[, , ]) <- exp(pars[83])
  mod_lmtxm_2fac
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxm_2fac <- fitSSM(mod_lmtxm_2fac, c(fit_lmtxm1$opt$par[1:41], -20:20/200, -6), updatefn5, method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_timestrt_time))
```

```
## [1] "calculation time is:1.0575065334638"
```

```
conf_lmtxm_2fac <- predict(fit_lmtxm_2fac$model, interval = "confidence", level=0.95)
pre_lmtxm_2fac <- predict(fit_lmtxm_2fac$model, interval = "prediction", level=0.95)
```

- 2因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 err_lmtxm_2fac を算出する。

```
obs_lmtxm_2fac <- lmtxm_nonNA

pred_lmtxm_2fac <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxm_2fac[nm] <- data.frame(pred_lmtxm_2fac[nm])[un_2011, 1]
}

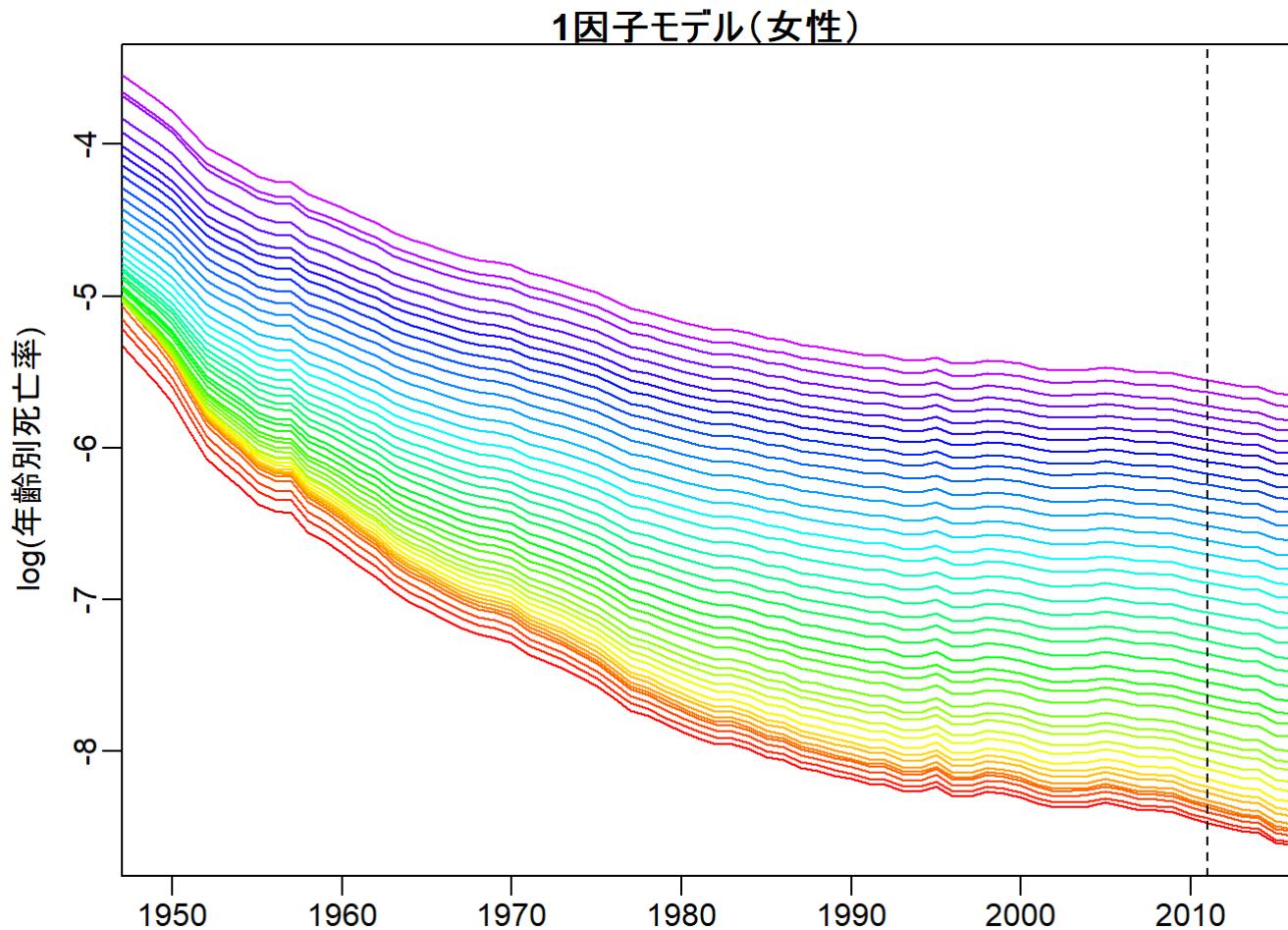
err_lmtxm_2fac <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxm_2fac[num_vec] <- pred_lmtxm_2fac[num_vec] - obs_lmtxm_2fac[num_vec]
}
```

解析結果の解釈

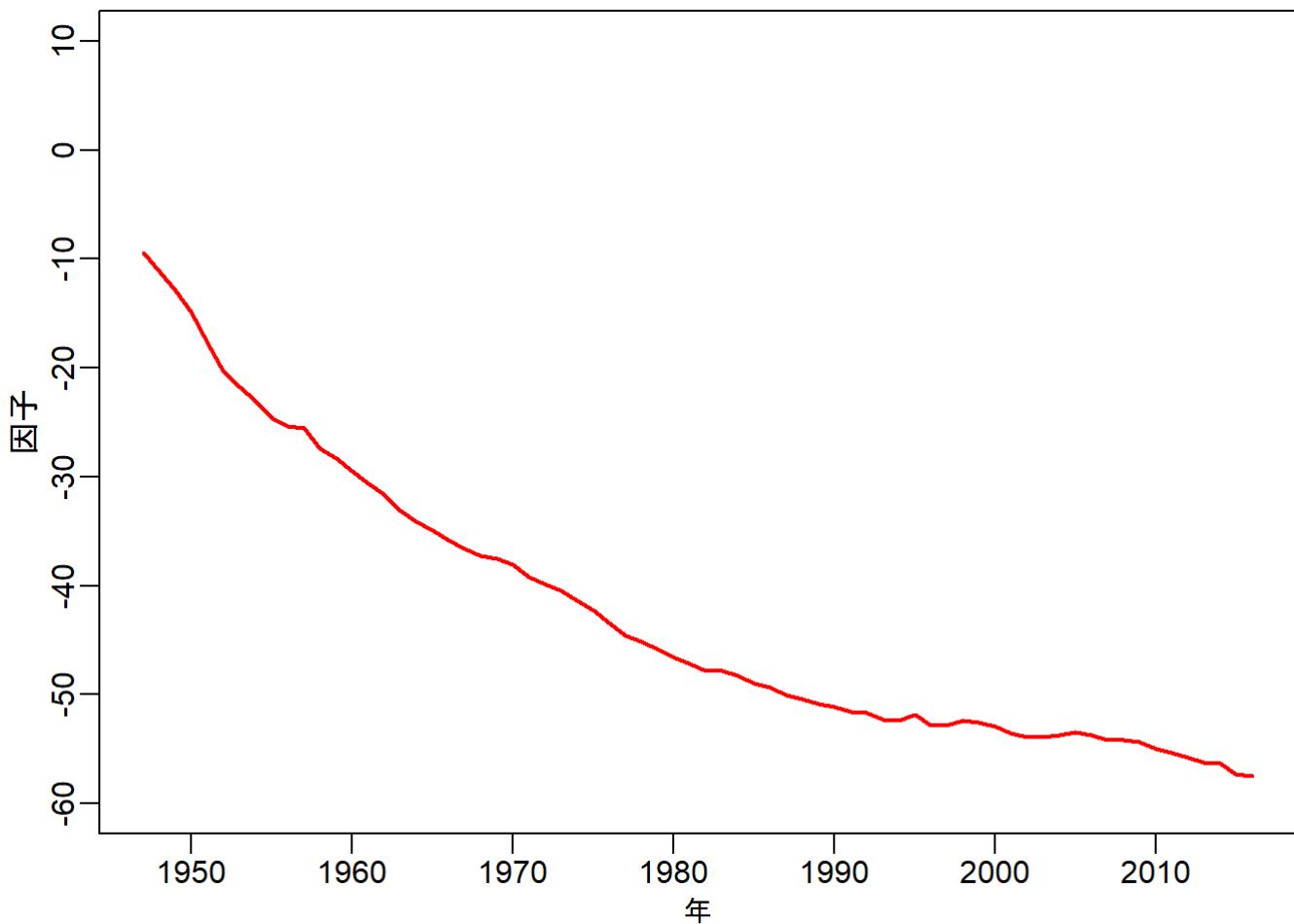
- 上記の解析で得た結果をもとに、2011年の20~60歳男女の年齢別死亡率について、予測値と観測値を比べてみる。
- まず、女性・男性それぞれにおけるモデル（1因子・2因子）の予測値を可視化する。

```
kfs_lmtxf <- KFS(fit_lmtxf1$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lmtxf$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log(年齢別死亡率)", main="1因子モデル（女性）")
abline(v=2011, lty=2)
```



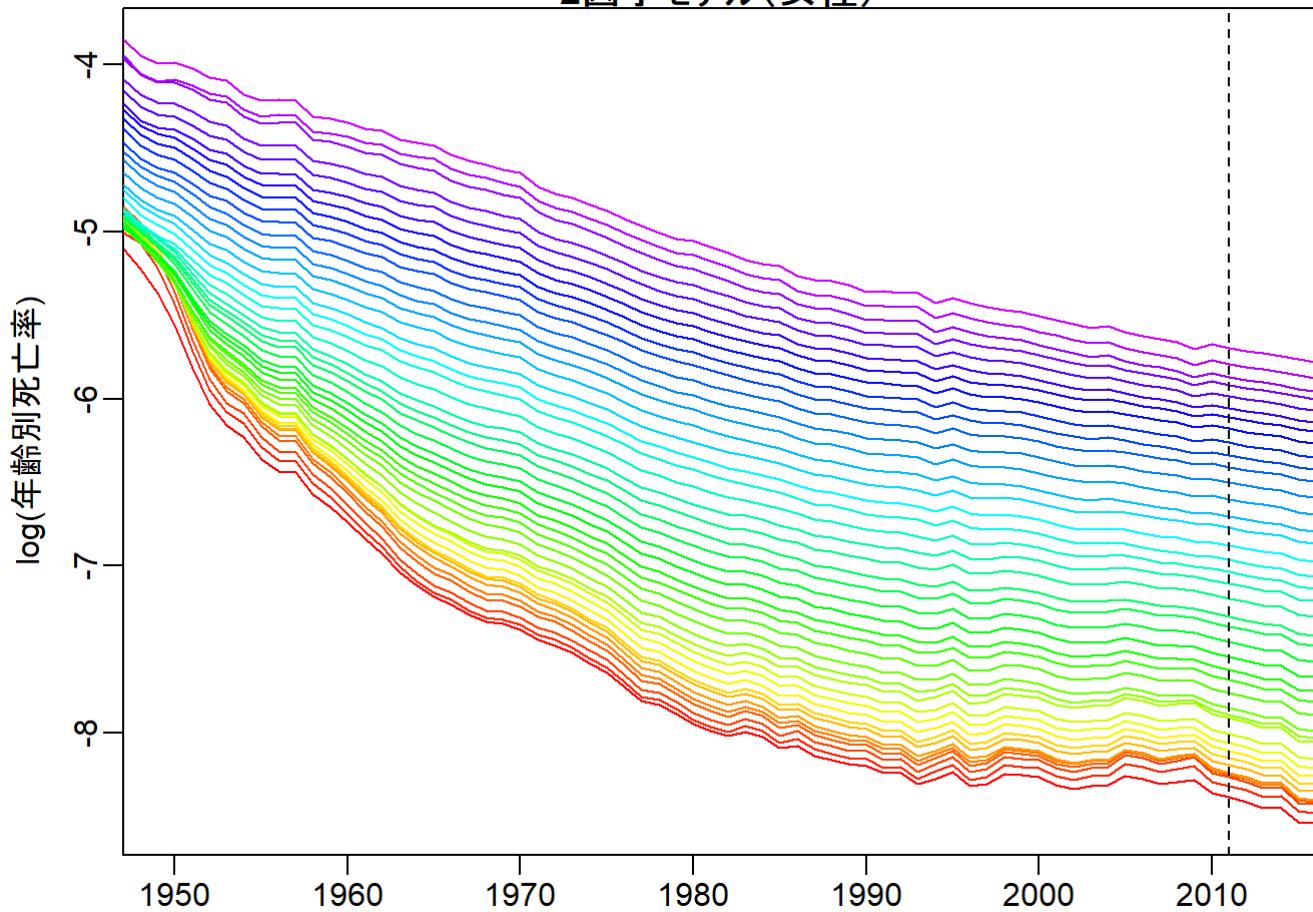
```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxf$alphahat[, 1], type="l", col=2, lwd=2, xlab="年", ylab="因子", ylim=c(-60, 10))
```



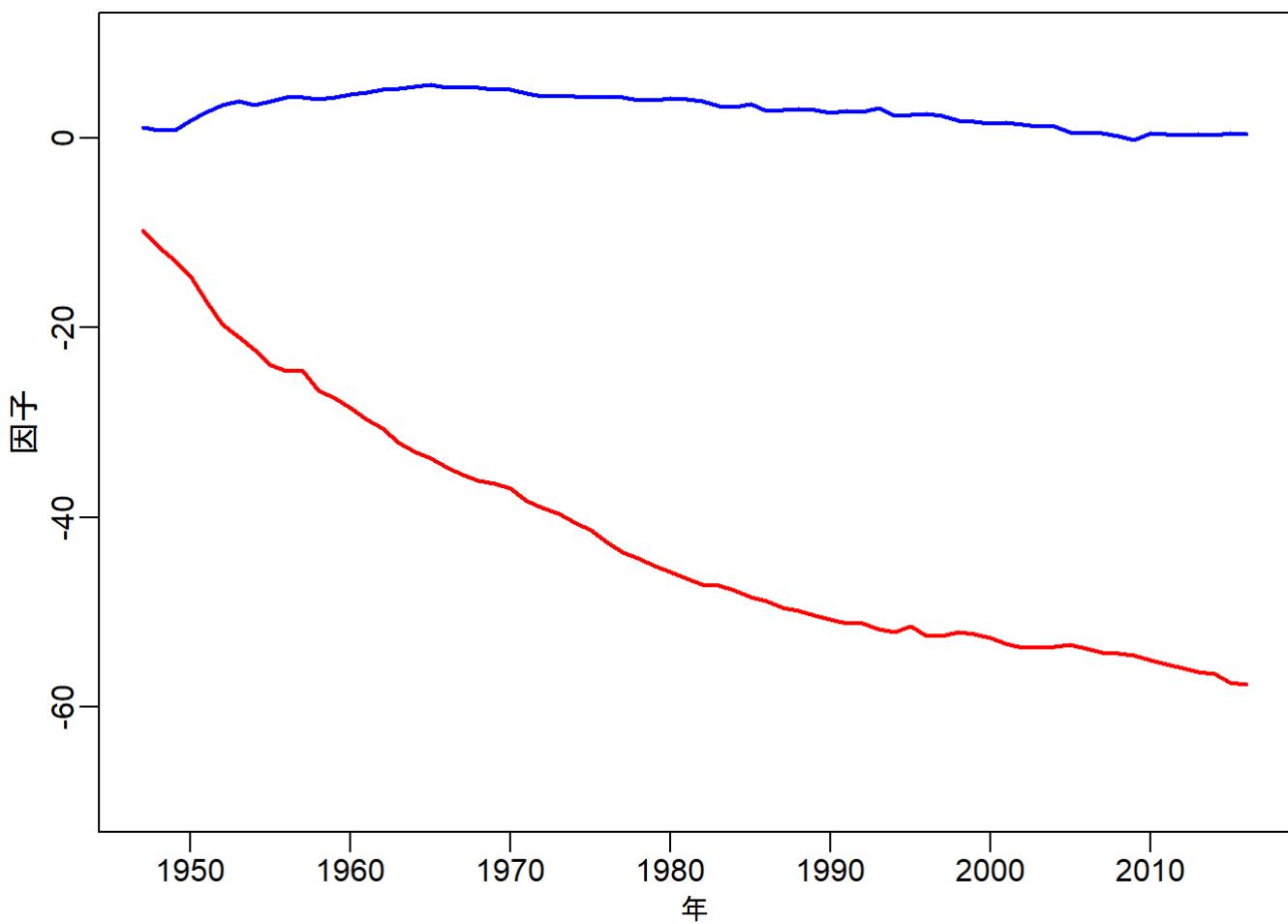
```
kfs_lmtxf_2fac <- KFS(fit_lmtxf_2fac$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lmtxf_2fac$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log
(年齢別死亡率)", main="2因子モデル (女性) ")
abline(v=2011, lty=2)
```

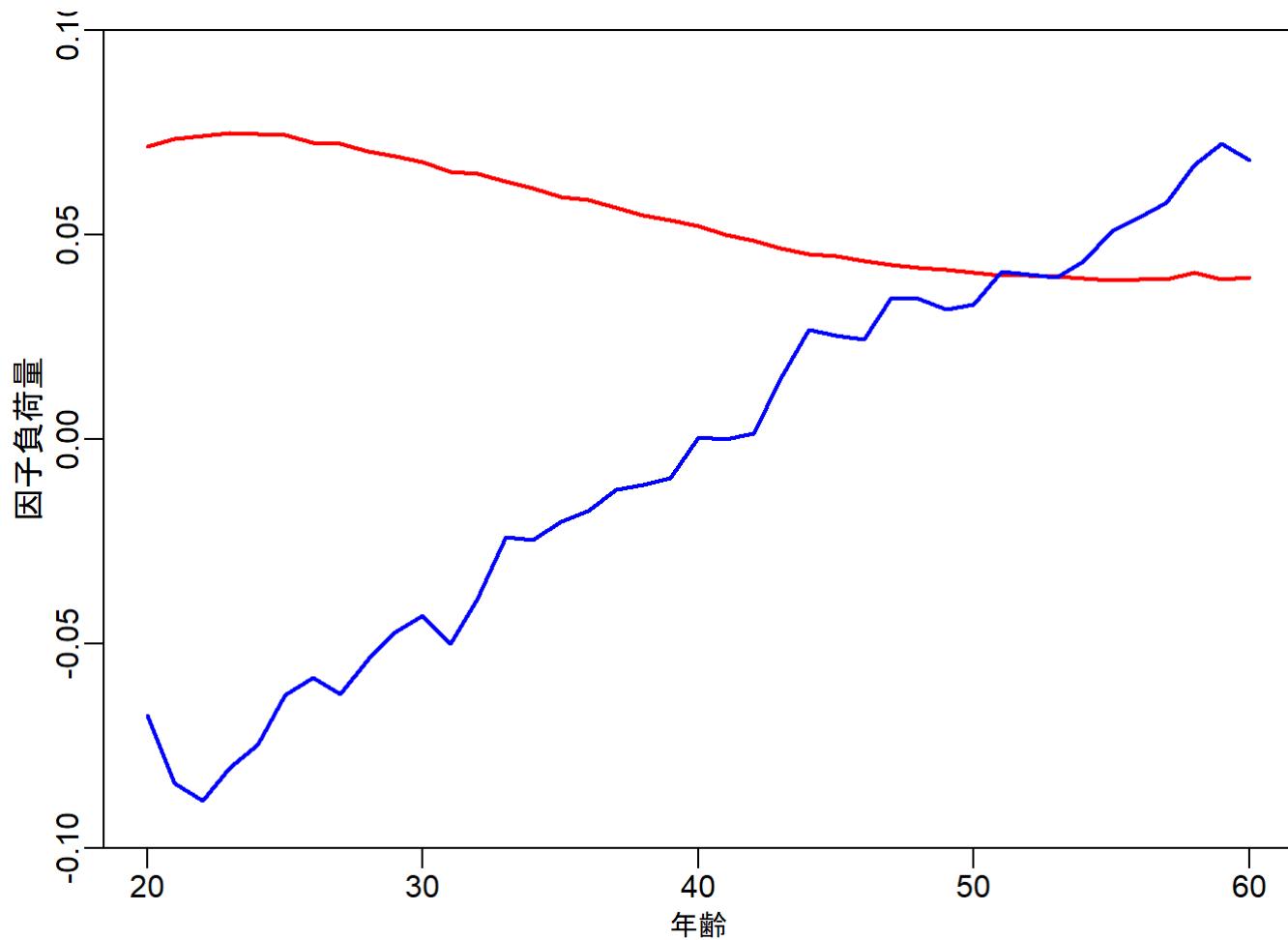
2因子モデル(女性)



```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxf_2fac$alphahat[, "factor1_level"], type="l", col=2, lwd=2, xlab="年", ylab="因子", ylim=c(-70, 10))
lines(1947:2016, kfs_lmtxf_2fac$alphahat[, "factor2_level"], type="l", col=4, lwd=2)
```



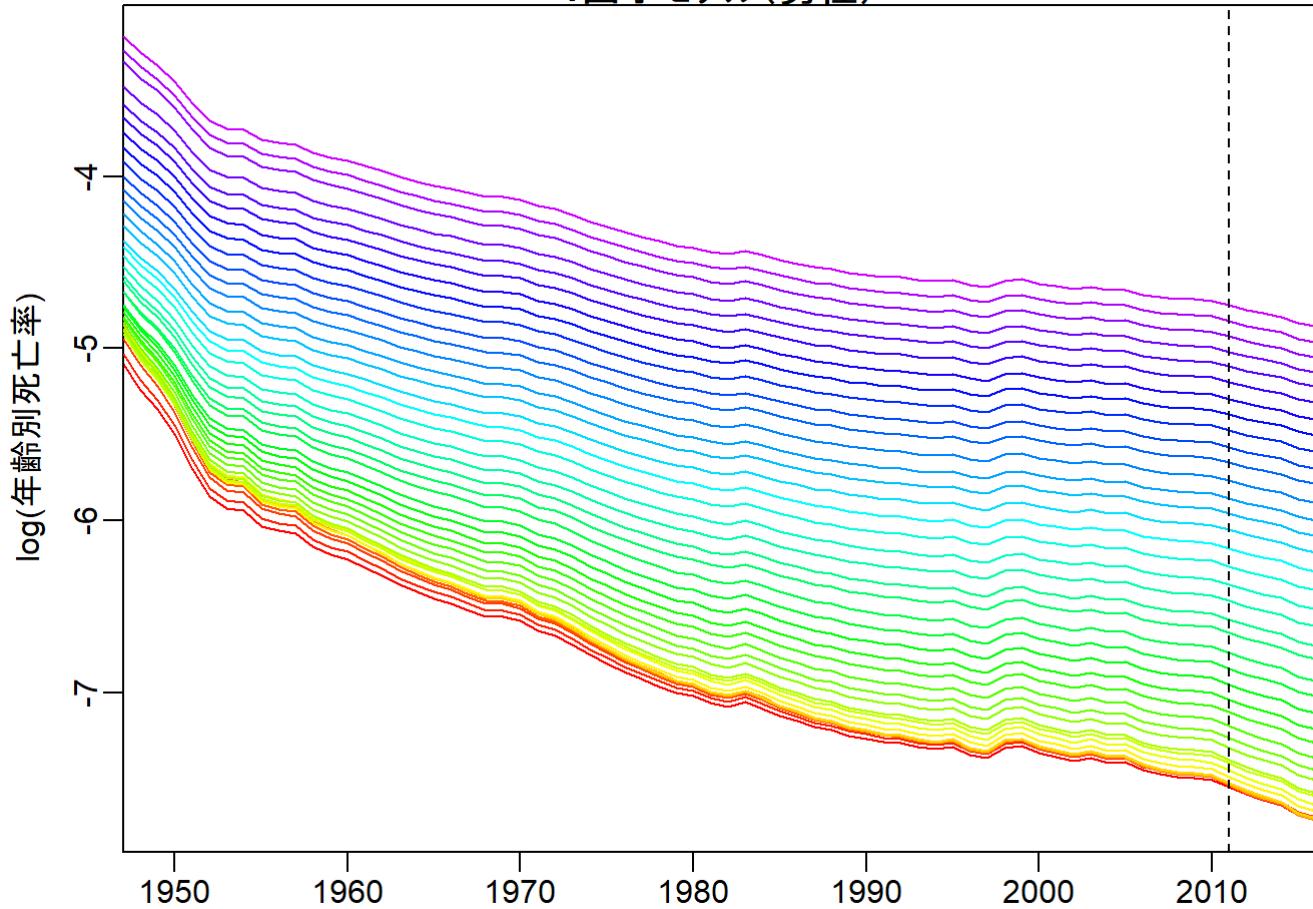
```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(20:60, fit_lmtxf_2fac$opt$par[1:41], type="l", col=2, lwd=2, xlab="年齢", ylab="因子負荷量", ylim=c(-0.1, 0.1), yaxs="i")
lines(20:60, fit_lmtxf_2fac$opt$par[42:82], col=4, lwd=2)
```



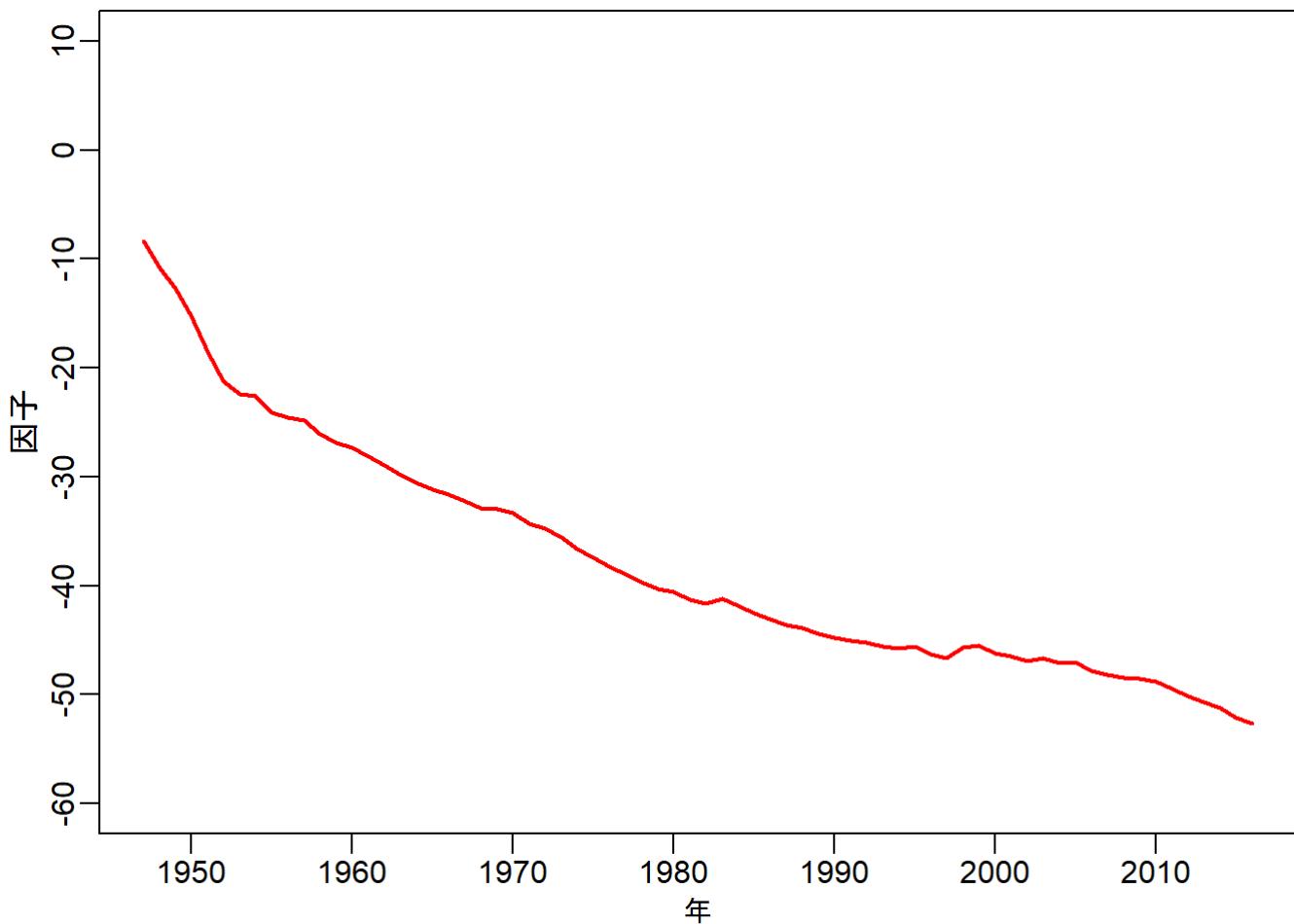
```
kfs_lmtxm <- KFS(fit_lmtxm1$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lmtxm$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log(年齢別死亡率)", main="1因子モデル（男性）")
abline(v=2011, lty=2)
```

1因子モデル(男性)



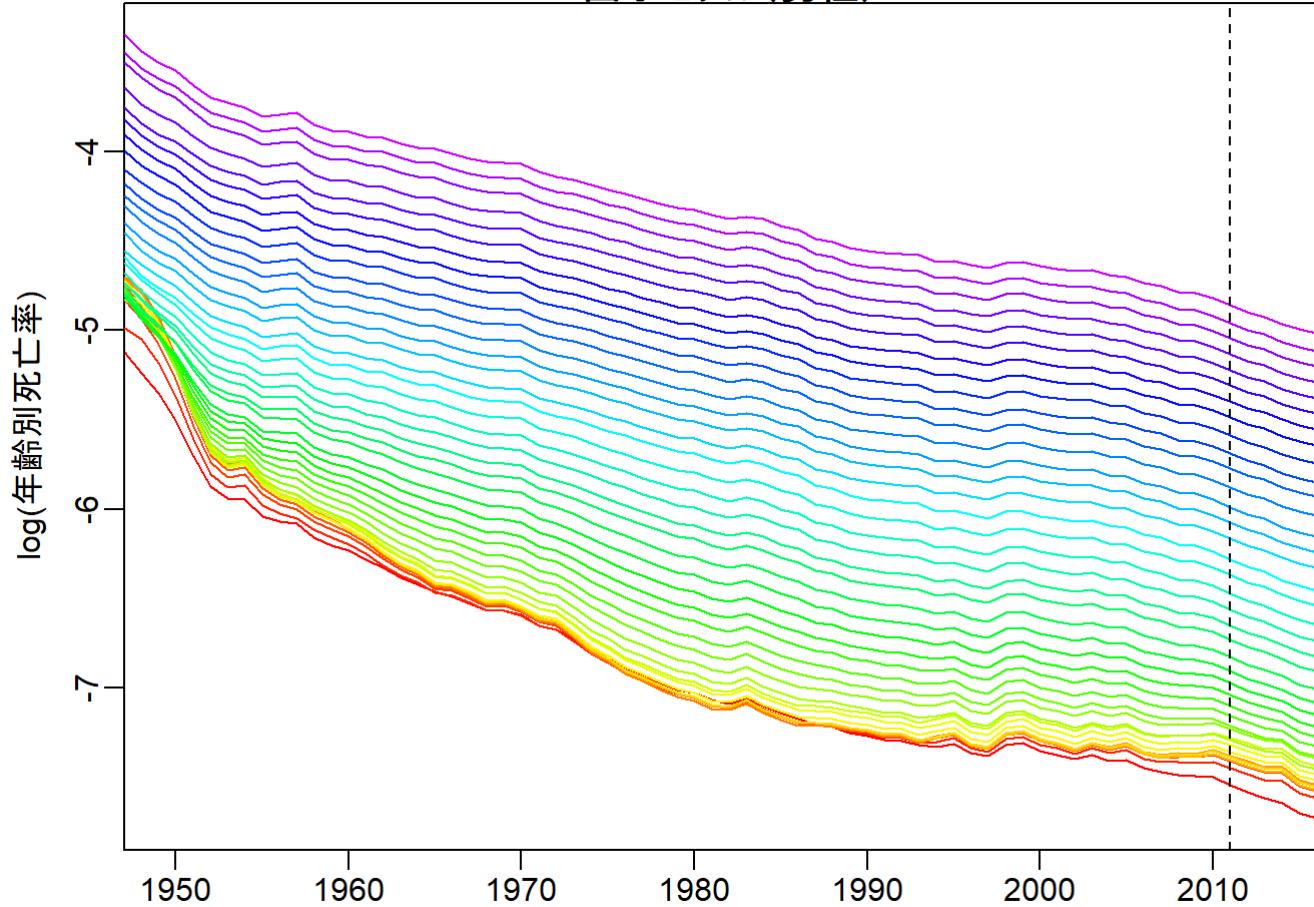
```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxm$alphahat[, 1], type="l", col=2, lwd=2, xlab="年", ylab="因子", ylim=c(-60, 10))
))
```



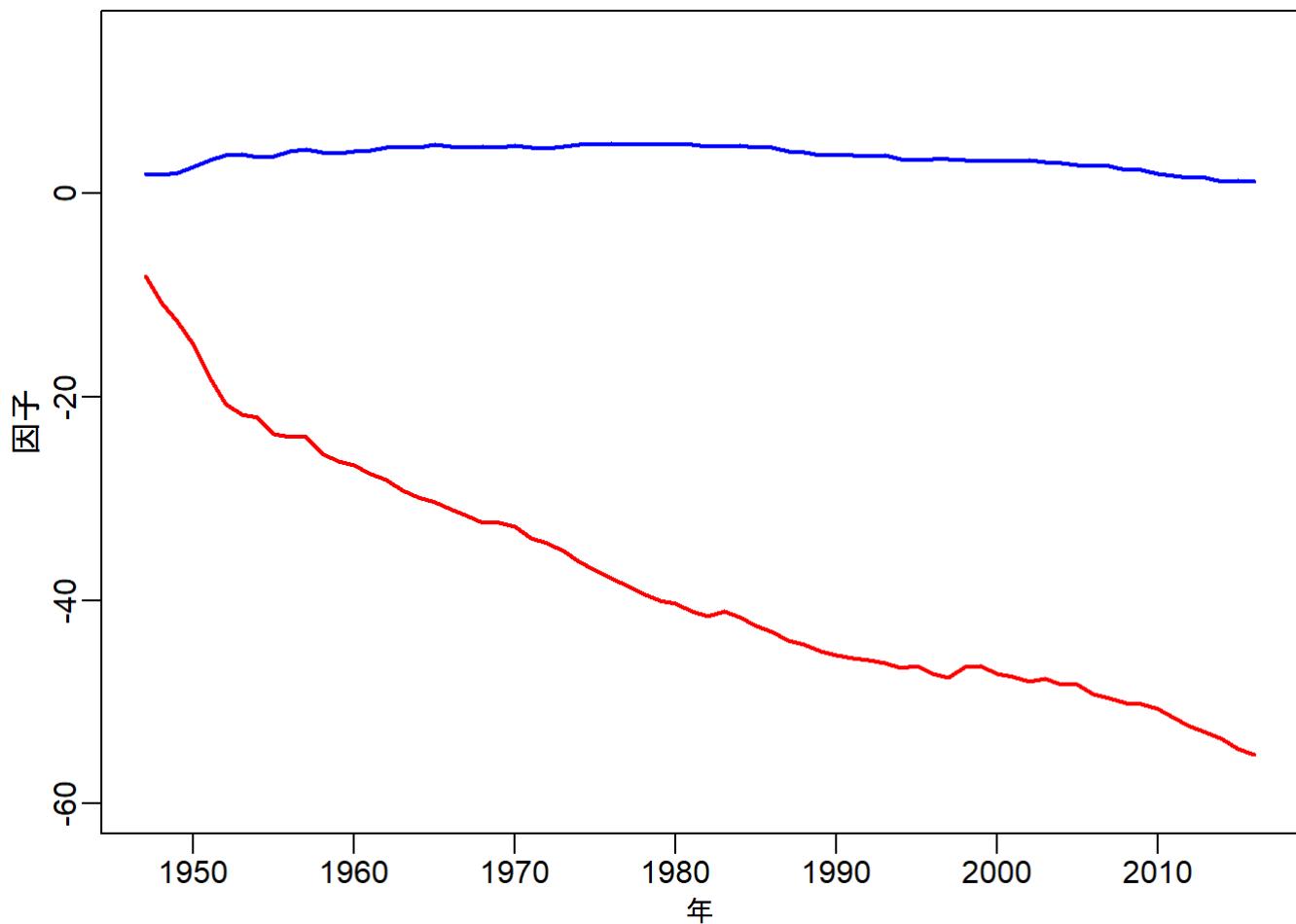
```
kfs_lmtxm_2fac <- KFS(fit_lmtxm_2fac$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lmtxm_2fac$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log
(年齢別死亡率)", main="2因子モデル (男性)")
abline(v=2011, lty=2)
```

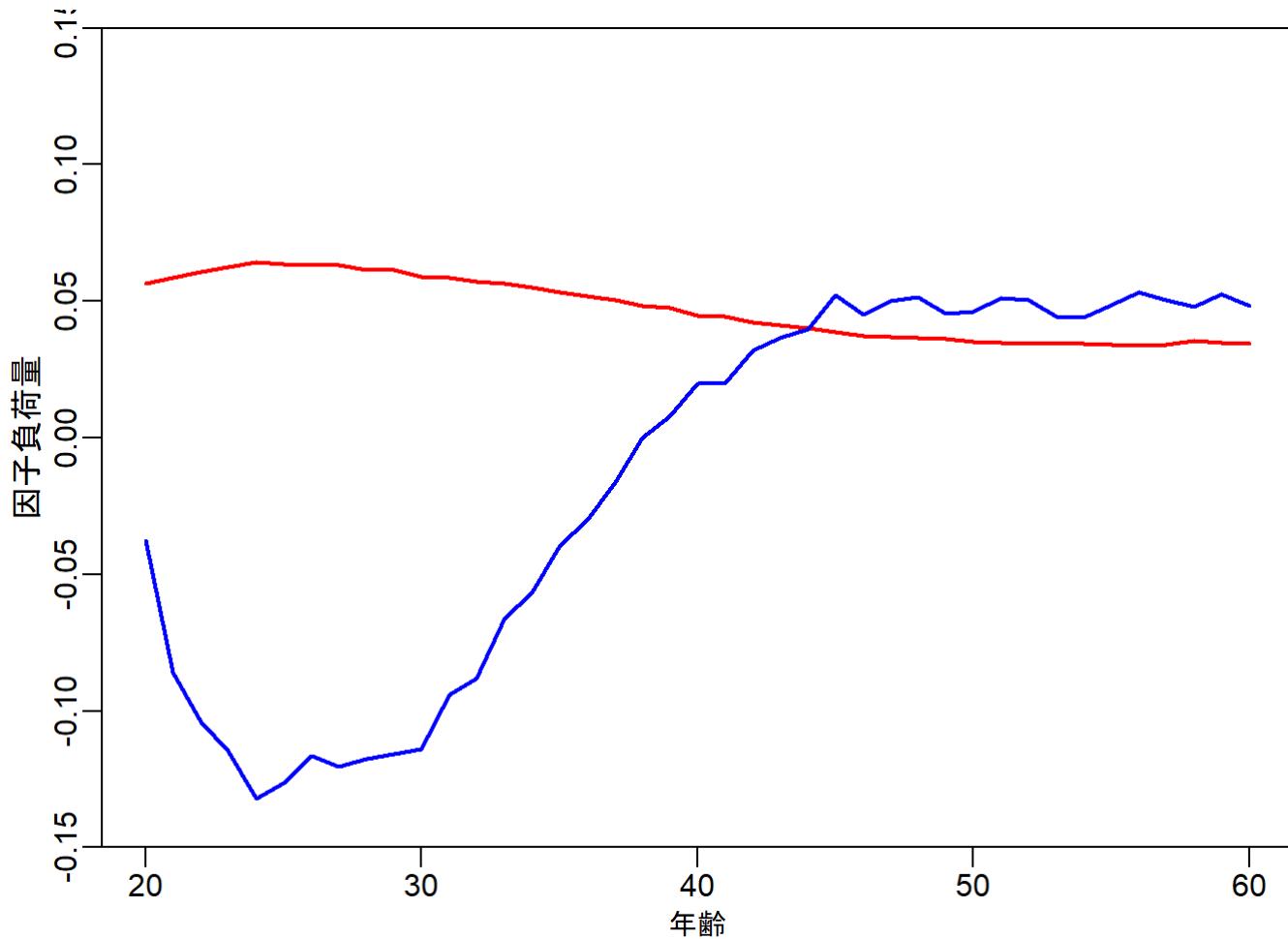
2因子モデル(男性)



```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxm_2fac$alphahat[, "factor1_level"], type="l", col=2, lwd=2, xlab="年", ylab="因子", ylim=c(-60, 15))
lines(1947:2016, kfs_lmtxm_2fac$alphahat[, "factor2_level"], type="l", col=4, lwd=2)
```



```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(20:60, fit_lmtxm_2fac$opt$par[1:41], type="l", col=2, lwd=2, xlab="年齢", ylab="因子負荷量", ylim=c(-0.15, 0.15), yaxs="i")
lines(20:60, fit_lmtxm_2fac$opt$par[42:82], col=4, lwd=2)
```

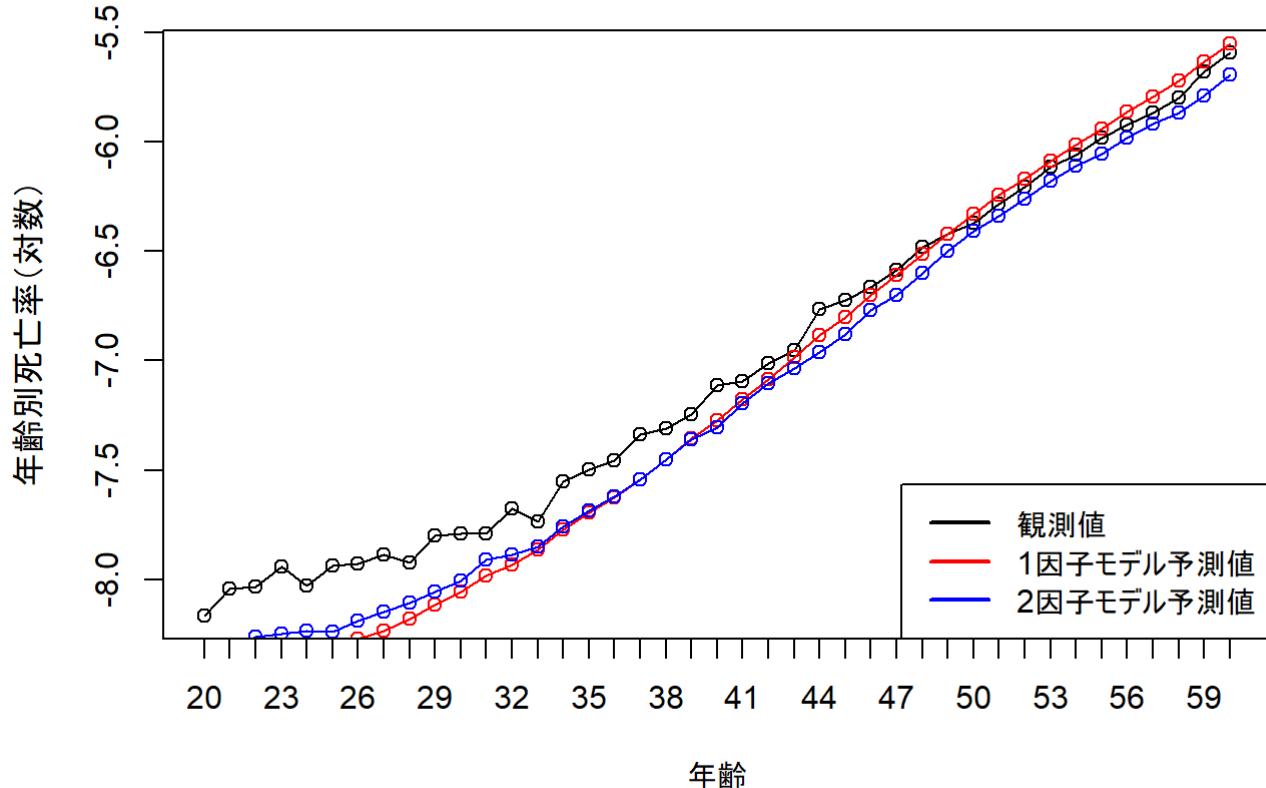


男性・2因子モデルの1950年前後における若い世代（20代前半）の死亡率が上がり、逆転現象が起きているが、これは第二次世界大戦による徴兵の影響を表現していると考えられる。- ちなみに、第1因子は時代、第2因子は年齢に対する因子を表している。

- 続いて、予測対象であった2011年の死亡率について、観測値とそれぞれのモデルでの予測値とを可視化し比較する。

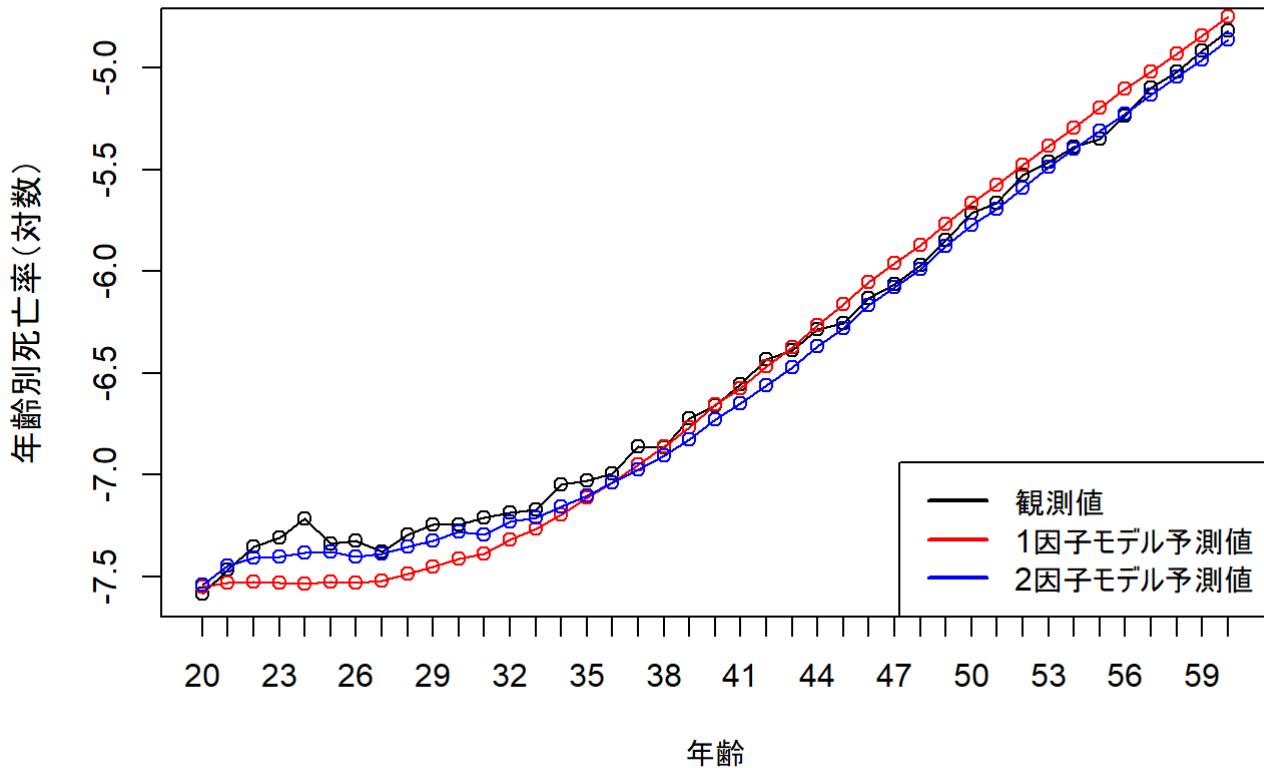
```
plot(obs_lmtxf, type="o", main="2011年の女性20-60歳死亡率（対数変換あり）", ylab="年齢別死亡率（対数）", xlab="年齢", xaxt="n")
lines(pred_lmtxf1, col="red", type="o")
lines(pred_lmtxf_2fac, col="blue", type="o")
legend("bottomright", c("観測値", "1因子モデル予測値", "2因子モデル予測値"), lwd=2, col=c(1, 2, 4))
axis(side=1, at=1:41, labels=c(thr))
```

2011年の女性20-60歳死亡率(対数変換あり)



```
plot(obs_lmtxm, type="o", main="2011年の女性20-60歳死亡率(対数変換あり)", ylab="年齢別死亡率(対数)", xlab="年齢", xaxt="n")
lines(pred_lmtxm1, col="red", type="o")
lines(pred_lmtxm_2fac, col="blue", type="o")
legend("bottomright", c("観測値", "1因子モデル予測値", "2因子モデル予測値"), lwd=2, col=c(1, 2, 4))
axis(side=1, at=1:41, labels=c(thr))
```

2011年の男性20-60歳死亡率(対数変換あり)

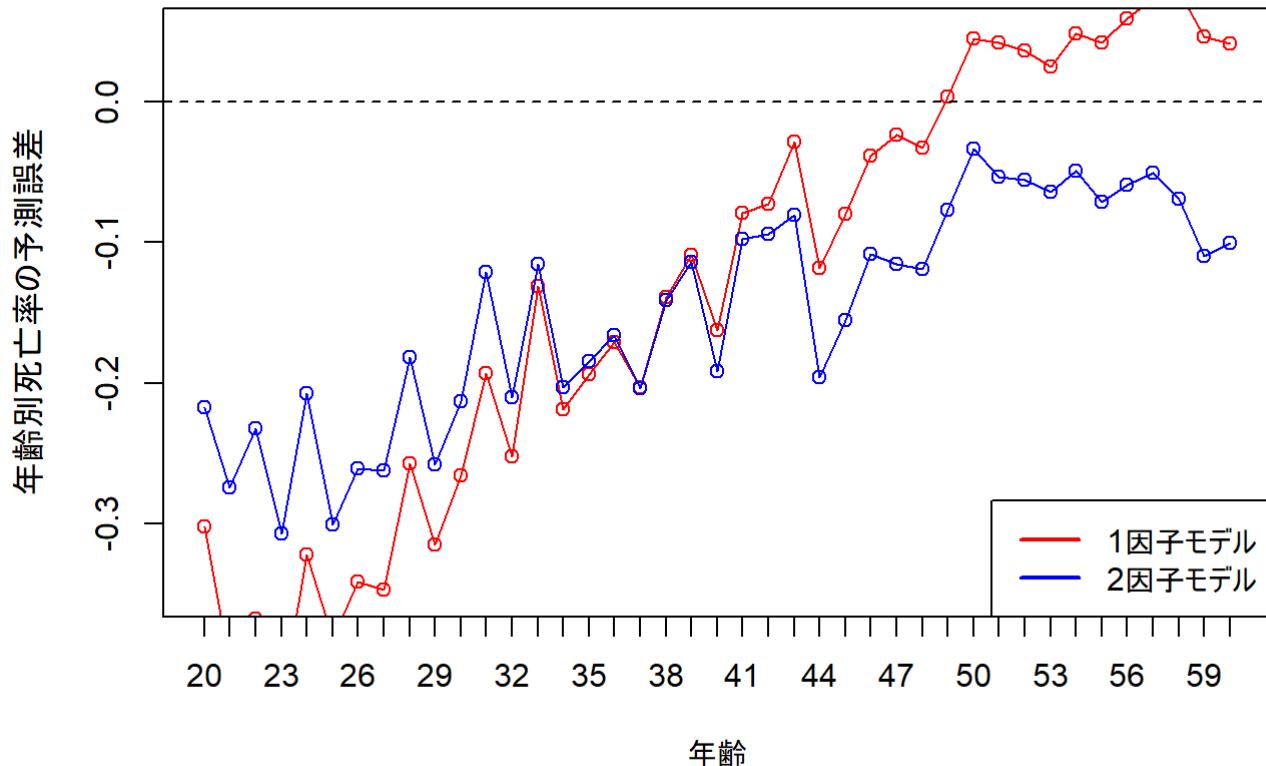


両者とも、若い世代では1因子モデルの方が若干観測値と近い値を予測できているように見える一方、年配の世代は2因子モデルの方が観測値と近い値を予測しているように見える。

- 次に、先ほど算出した、観測値と予測値の誤差について可視化する。

```
plot(err_lmtxf, type="o", main="2011年20-60歳死亡率予測誤差（女性）", ylab="年齢別死亡率の予測誤差", xlab="年齢", xaxt="n", col="red", ylim=c(-0.35, 0.05))
lines(err_lmtxf_2fac, col="blue", type="o")
legend("bottomright", c("1因子モデル", "2因子モデル"), lwd=2, col=c(2, 4))
axis(side=1, at=1:41, labels=c(thr))
abline(h=0, lty=2)
```

2011年20-60歳死亡率予測誤差(女性)

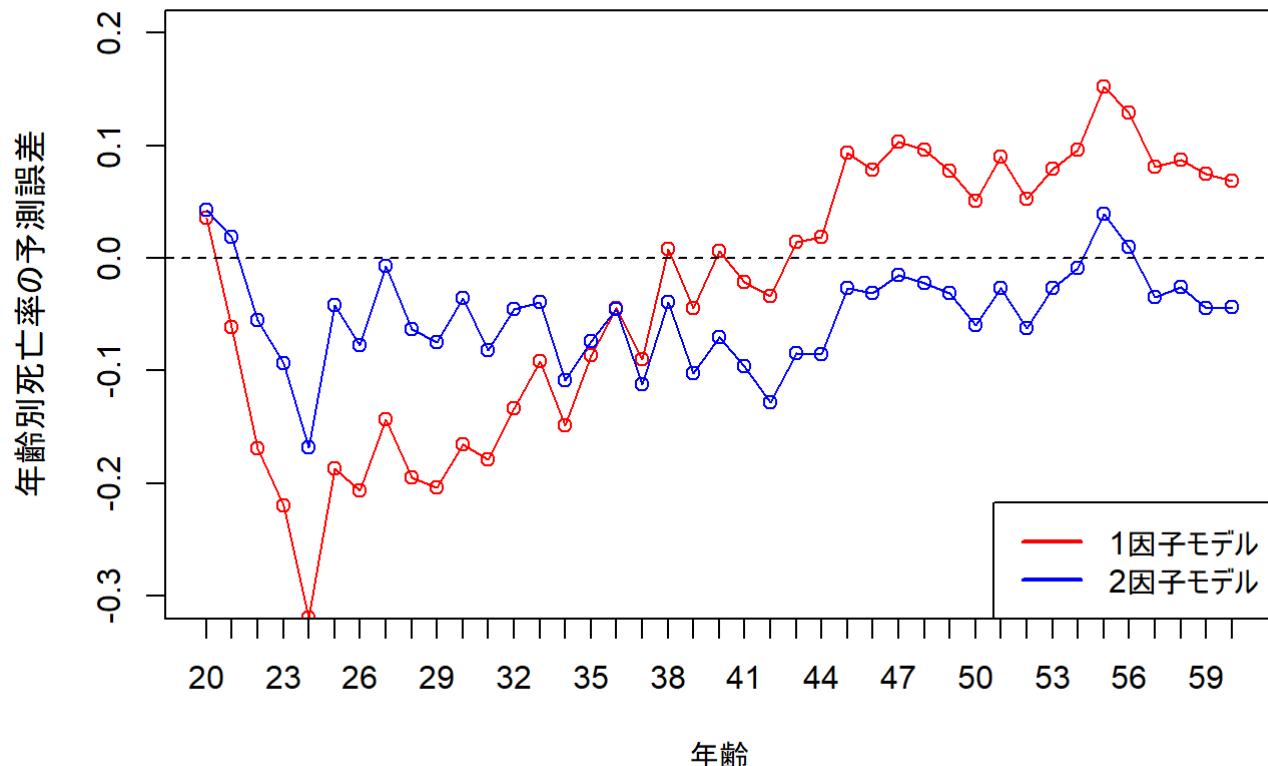


```

plot(err_lmtxm, type="o", main="2011年20-60歳死亡率予測誤差（男性）", ylab="年齢別死亡率の予測誤差", xlab="年齢", xaxt="n", col="red", ylim=c(-0.3, 0.2))
lines(err_lmtxm_2fac, col="blue", type="o")
legend("bottomright", c("1因子モデル", "2因子モデル"), lwd=2, col=c(2, 4))
axis(side=1, at=1:41, labels=c(thr))
abline(h=0, lty=2)

```

2011年20-60歳死亡率予測誤差(男性)

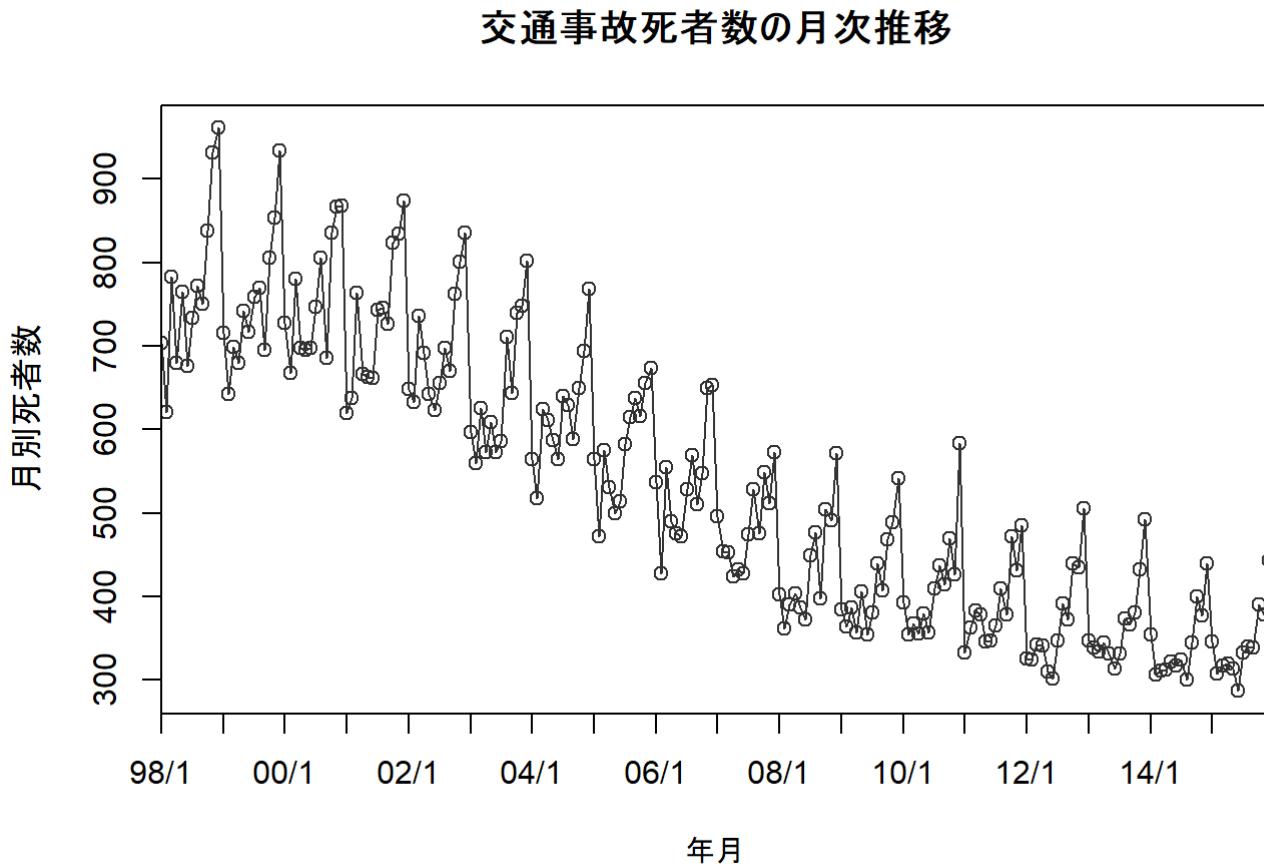


両者とも、若い世代では1因子モデルの方が若干予測誤差が小さい一方、年配の世代は2因子モデルの方が若干予測誤差が小さい。 - これは、20代の若い世代間での死亡率にあまり差がなく、戦争や災害といった時代のイベントによる変動の影響を受けやすい一方、高齢になってくると、そういうイベントだけでなく、年齢による死亡率の変動も比較的大きくなり、年齢の影響を受けやすくなるためと考えられる。

課題3

選択課題③ 解析例④の交通事故死者数の月次推移について観測値が正規分布に従う通常の構造時系列モデルを適用し、より良いモデルを追求する。

```
plot(tsuki, type="o", main="交通事故死者数の月次推移", xaxt="n", ylab="月別死者数", xlab="年月", col=gray(0.2), xaxs="i")
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```



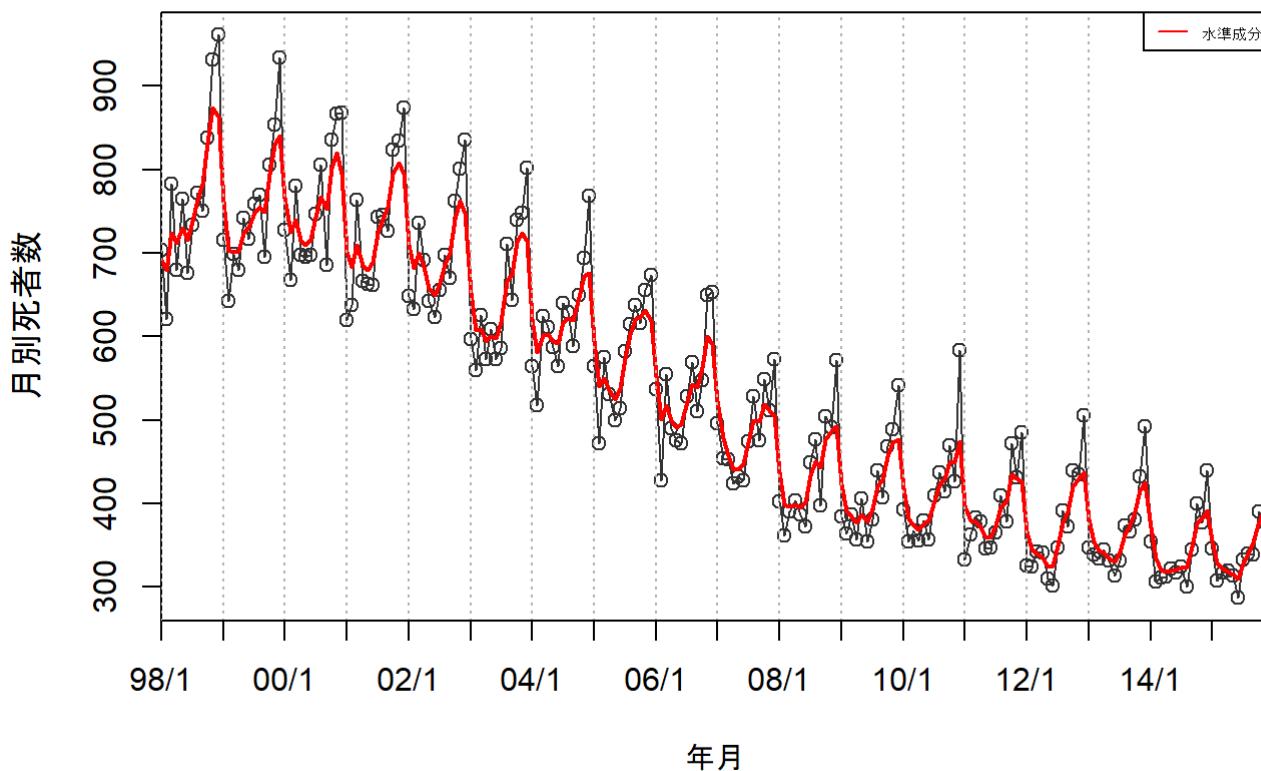
- モデル1：
 - まずはローカルレベルモデルで解析し、モデルの当てはまりを評価するためにAICを算出する。

```
# ローカルレベルモデル
mod_tri <- SSMModel(tsuki ~ SSMtrend(1, Q=NA), H=NA)
fit_tri <- fitSSM(mod_tri, numeric(2), method = "SANN")
kfs_tri <- KFS(fit_tri$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri <- kfs_tri$logLik - sum(kfs_tri$Finf>0) * log(2*pi)/2
AIC_tri <- -2*logLik_tri + 2*2
print(paste("AIC_ローカルレベルモデル : ", AIC_tri))
```

```
## [1] "AIC_ローカルレベルモデル : 2460.76641910979"
```

```
plot(tsuki, type="o", main="ローカルレベルモデル", xaxt="n", ylab="月別死者数", xlab="年月", col=gray(0.2), xaxs="i")
lines(kfs_tri$alphahat[, "level"], col=2, lwd=2)
legend("topright", "水準成分", lwd=1, col=2, cex = 0.5)
abline(v=1+0:15*12, lty=3, col=gray(0.7))
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

ローカルレベルモデル



- モデル2：
 - データのプロットを見ると、1年単位の周期性のような変動傾向が見られるので、先ほどのローカルレベルモデルに季節変動を加えてみる。

```
# ローカルレベルモデル+固定季節変動（対数変換なし）
mod_tri1 <- SSMModel(tsuki ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=0 ), H=NA)
fit_tri1 <- fitSSM(mod_tri1, numeric(2), method = "SANN")
kfs_tri1 <- KFS(fit_tri1$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri1 <- kfs_tri1$logLik - sum(kfs_tri1$Finf>0) * log(2*pi)/2
AIC_tri1 <- -2*logLik_tri1 + 2*( 2 + 12 )
print(paste("AIC_ローカルレベルモデル+固定季節変動 : ", AIC_tri1))
```

```
## [1] "AIC_ローカルレベルモデル+固定季節変動 : 2078.57392707883"
```

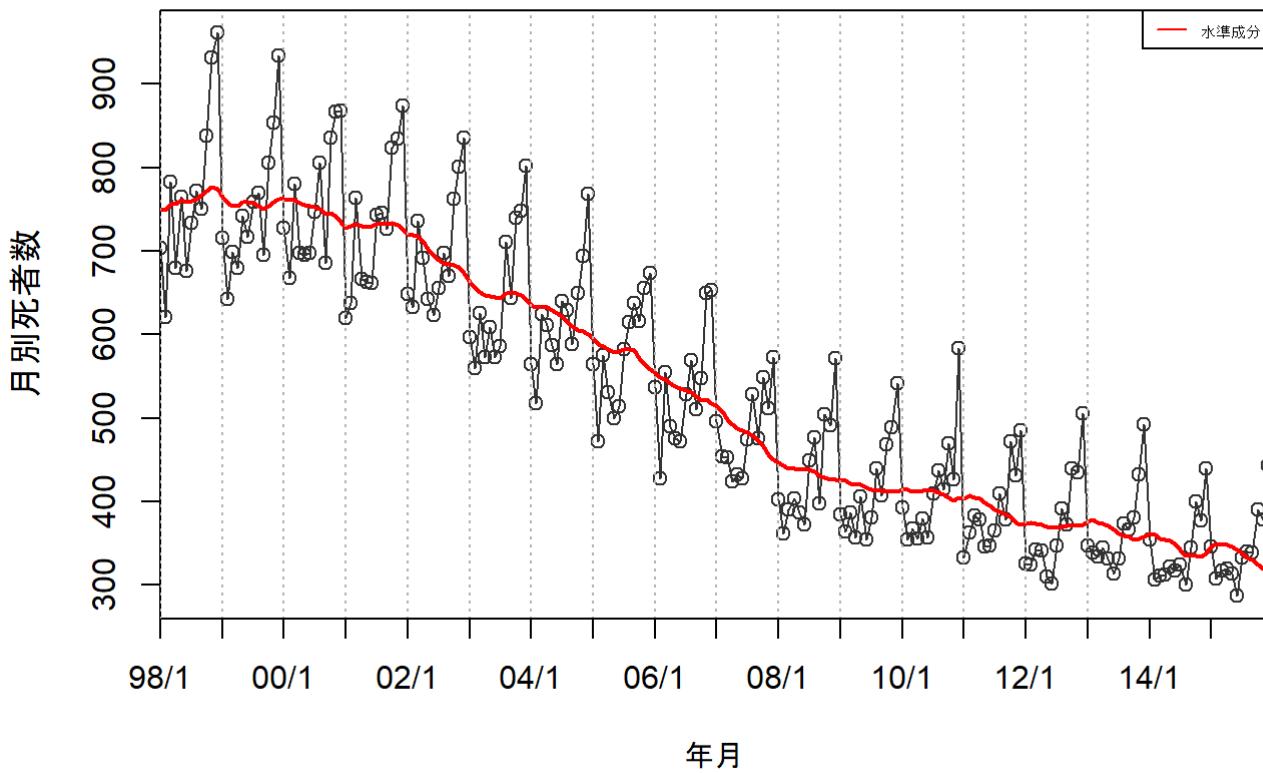
```

plot(tsuki, type="o", main="ローカルレベルモデル+固定季節変動", xaxt="n", ylab="月別死者数", xlab=
"年月", col=gray(0.2), xaxs="i")
lines(kfs_tri1$alphahat[, "level"], col=2, lwd=2)

legend("topright", "水準成分", lwd=1, col=2, cex = 0.5)
abline(v=1+0:15*12, lty=3, col=gray(0.7))
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))

```

ローカルレベルモデル+固定季節変動



- モデル3：
 - 季節変動の程度がありそなうなので、固定季節変動を可変季節変動に変更してみる。

```

# ローカルレベルモデル+可変季節変動（対数変換なし）
mod_tri2 <- SSModel(tsuki ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=NA), H=NA)
fit_tri2 <- fitSSM(mod_tri2, numeric(3), method = "SANN")
kfs_tri2 <- KFS(fit_tri2$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri2 <- kfs_tri2$logLik - sum(kfs_tri2$Finf>0) * log(2*pi)/2
AIC_tri2 <- -2*logLik_tri2 + 2*( 3 + 12 )
print(paste("AIC_ローカルレベルモデル+可変季節変動 : ", AIC_tri2))

```

```
## [1] "AIC_ローカルレベルモデル+可変季節変動 : 2043.09870233152"
```

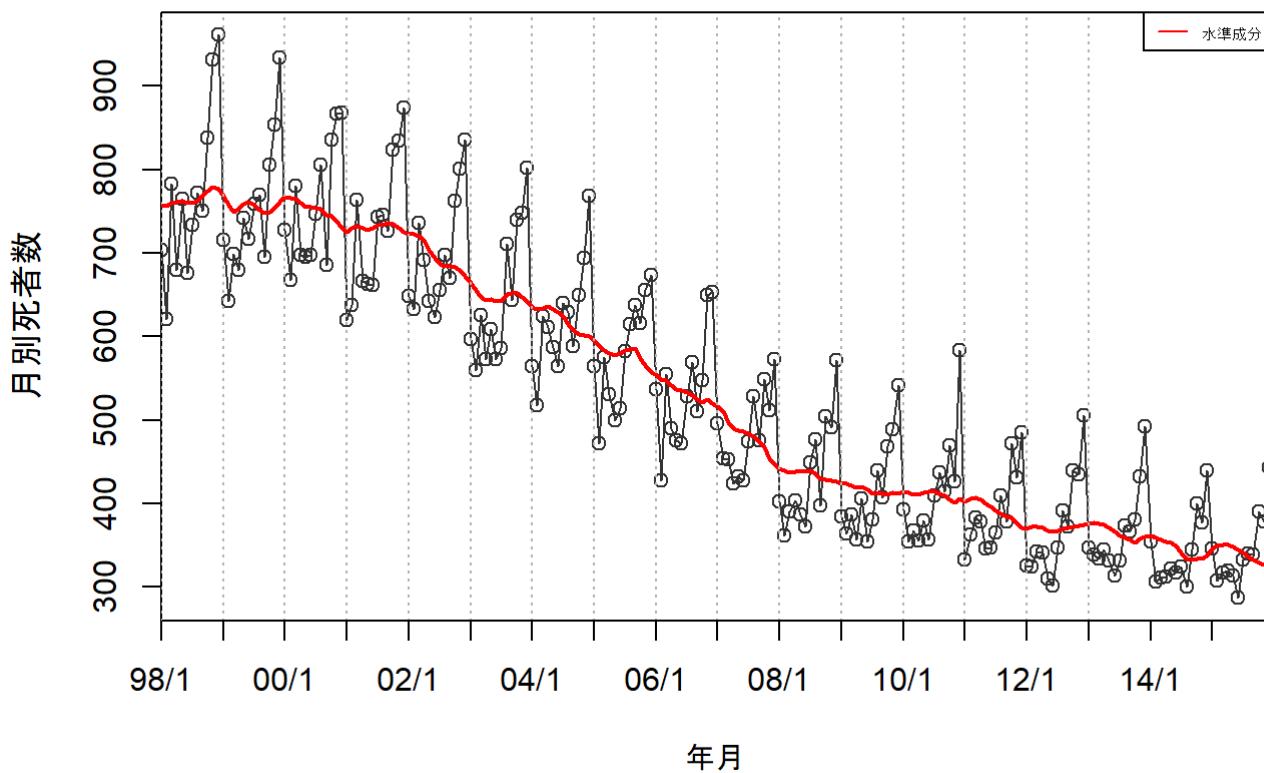
```

plot(tsuki, type="o", main="ローカルレベルモデル+可変季節変動", xaxt="n", ylab="月別死者数", xlab=
"年月", col=gray(0.2), xaxs="i")
lines(kfs_tri2$alphahat[, "level"], col=2, lwd=2)

legend("topright", "水準成分", lwd=1, col=2, cex = 0.5)
abline(v=1+0:15*12, lty=3, col=gray(0.7))
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))

```

ローカルレベルモデル+可変季節変動



- モデル4:
 - モデル3を平滑化トレンドモデルにしてみる。

```

# 平滑化トレンドモデル+可変季節変動（対数変換なし）
mod_tri3 <- SSMModel(tsuki ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit_tri3 <- fitSSM(mod_tri3, numeric(3), method = "SANN")
kfs_tri3 <- KFS(fit_tri3$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri3 <- kfs_tri3$logLik - sum(kfs_tri3$Finf > 0) * log(2*pi)/2
AIC_tri3 <- -2*logLik_tri3 + 2*( 3 + 13 )
print(paste("AIC_平滑化トレンドモデル+可変季節変動 : ", AIC_tri3))

```

```
## [1] "AIC_平滑化トレンドモデル+可変季節変動 : 2032.90509293277"
```

- ここまでで最も良いAICを記録したのは、平滑化トレンドモデル+可変季節変動。
- 続いて、上記モデルで目的変数に対数変換を施してみる。

```
# 平滑化トレンドモデル+可変季節変動（対数変換あり）
mod_tri4 <- SSModel(log(tsuki) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit_tri4 <- fitSSM(mod_tri4, numeric(3), method = "SANN")
kfs_tri4 <- KFS(fit_tri4$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri4 <- kfs_tri4$logLik - sum(kfs_tri4$Finf > 0) * log(2*pi)/2
AIC_tri4 <- -2*logLik_tri4 + 2*( 3 + 13 ) + 2*sum(log(tsuki))
print(paste("AIC_平滑化トレンドモデル+可変季節変動（対数変換あり） : ", AIC_tri4))
```

```
## [1] "AIC_平滑化トレンドモデル+可変季節変動（対数変換あり） : 2174.93509205638"
```

- AICが悪化してしまった。
- もう少し違うパターンからより良いモデルを探索してみることにする。
- 観測値のプロットに立ち戻ると、およそ1年単位での変動周期が、毎年上下動している。
- そこで、次のモデルを定義する。
- モデル6:
 - 平滑化トレンドモデル+可変季節変動+前年同期比（対数変換あり）

```
# 平滑化トレンドモデル+可変季節変動（対数変換あり）
lag_tri6 <- c(tsuki/stats::lag(tsuki, k=-12))
first <- rep(1, length=12)
lag_tri6 <- ts	append(first, lag_tri6))
mod_tri6 <- SSModel(log(tsuki) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA) + SSMregression(log(tsuki) ~ lag_tri6, Q = NA), H = NA)
fit_tri6 <- fitSSM(mod_tri6, numeric(4), method = "SANN")
kfs_tri6 <- KFS(fit_tri6$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri6 <- kfs_tri6$logLik - sum(kfs_tri6$Finf > 0) * log(2*pi)/2
AIC_tri6 <- -2*logLik_tri6 + 2*( 4 + 12 ) + 2*sum(log(tsuki))
print(paste("AIC_平滑化トレンドモデル+可変季節変動+前年同期比 : ", AIC_tri6))
```

```
## [1] "AIC_平滑化トレンドモデル+可変季節変動+前年同期比 : 2018.27897295501"
```

- AICが若干向上した。
- 次に、「良いモデル」の追求について、「モデルの当てはまり」の観点から最良を示したモデル6が長期予測の観点からはどうなのがかを確かめてみる。
- 今回は、直近4年分を欠損値として予測対象にする。

```
tsukiNA <- tsuki
tsukiNA[157:216] <- NA

mod_tri_pre1 <- SSModel(log(tsukiNA) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA) + SSMregression(log(tsukiNA) ~ lag_tri6, Q = NA), H = NA)
fit_tri_pre1 <- fitSSM(mod_tri_pre1, numeric(4), method = "SANN")
conf_tri_pre1 <- predict(fit_tri_pre1$model, interval="confidence", level=0.95)
pre_tri_pre1 <- predict(fit_tri_pre1$model, interval="prediction", level=0.95)
```

- 予測結果をプロットしてみる。

```

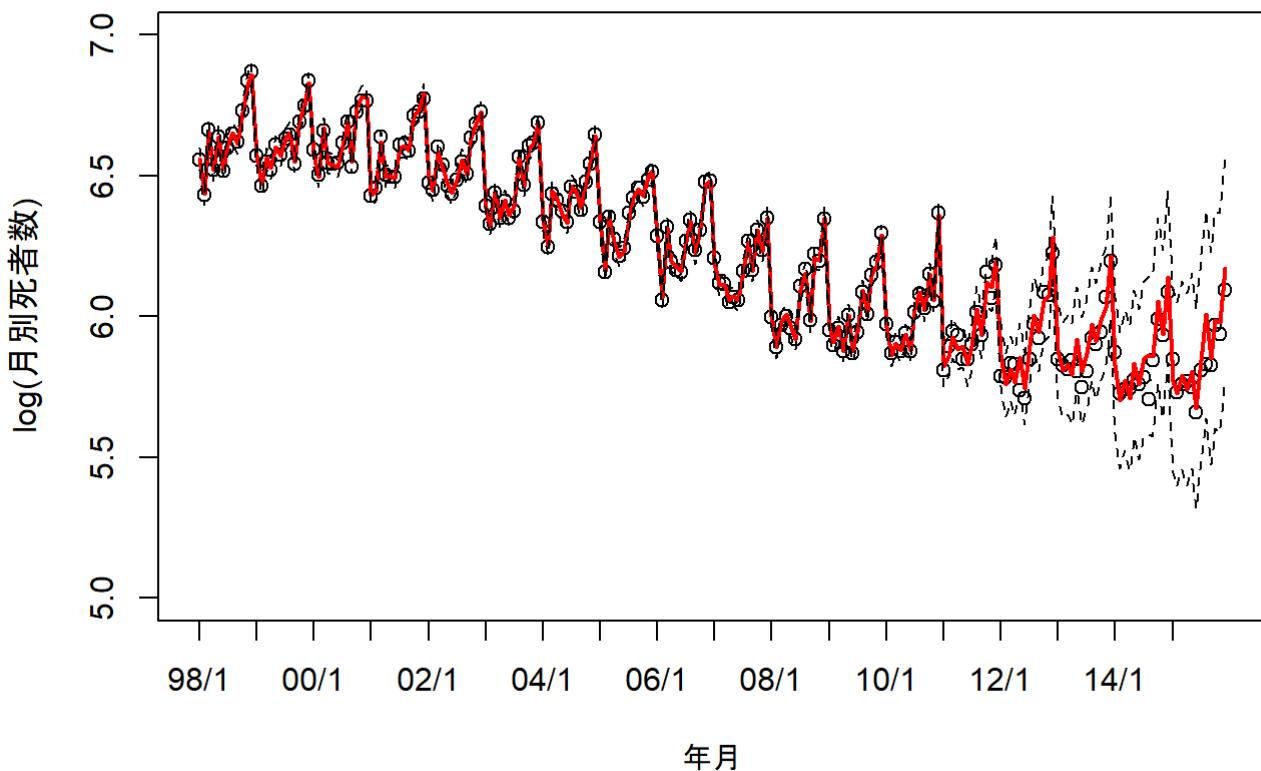
plot(log(tsuki), type="b", xaxt="n", main=" (最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比", ylab="log(月別死者数)", xlab="年月", lty=3, ylim=c(5, 7))

lines(pre_tri_pre1[, 1], lwd=2, col="red")
lines(pre_tri_pre1[, 2], lty=2)
lines(pre_tri_pre1[, 3], lty=2)

axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))

```

(最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比



- 比較のため、モデル4でも予測してみる。

```

mod_tri_pre2 <- SSModel(tsukiNA ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA ), H=NA)
fit_tri_pre2 <- fitSSM(mod_tri_pre2, numeric(3), method = "SANN")
conf_tri_pre2 <- predict(fit_tri_pre2$model, interval="confidence", level=0.95)
pre_tri_pre2 <- predict(fit_tri_pre2$model, interval="prediction", level=0.95)

```

- 予測結果をプロットしてみる。

```

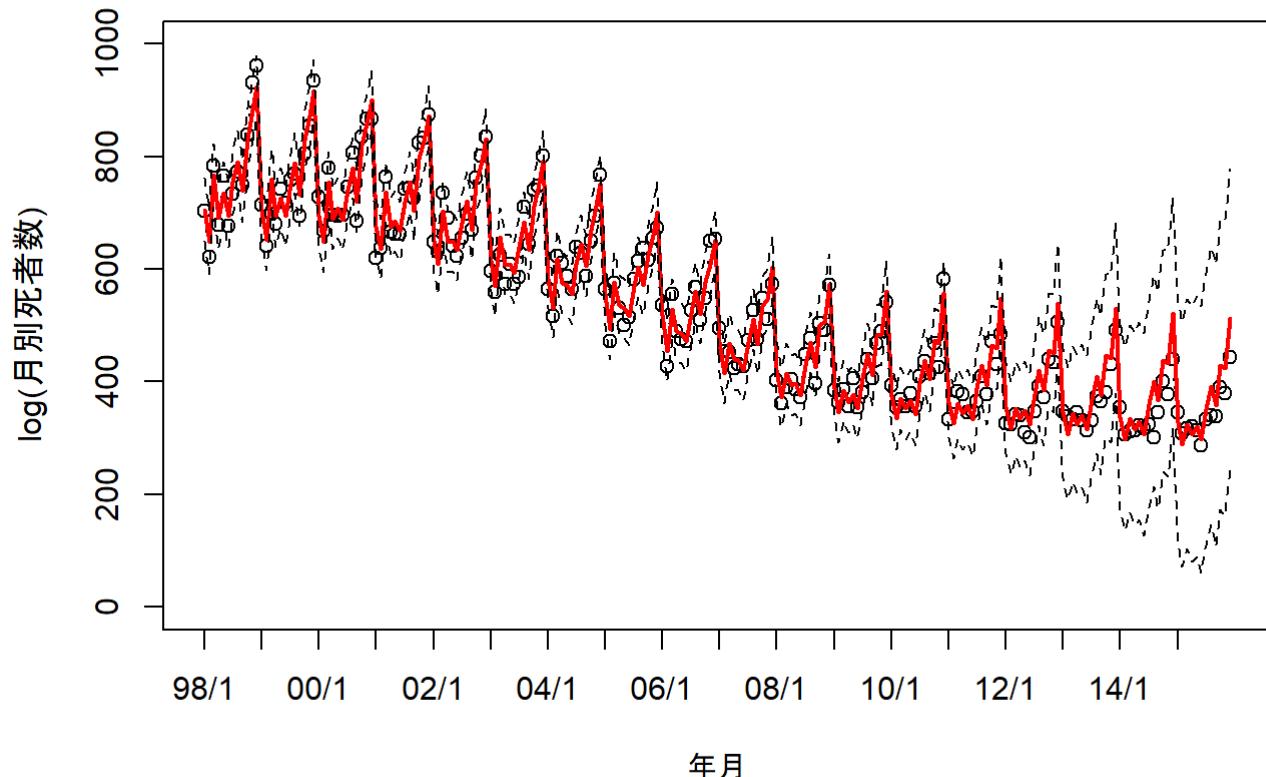
plot(tsuki, type="b", xaxt="n", main=" (最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比", ylab="log(月別死者数)", xlab="年月", lty=3, ylim=c(0, 1000))

lines(pre_tri_pre2[, 1], lwd=2, col="red")
lines(pre_tri_pre2[, 2], lty=2)
lines(pre_tri_pre2[, 3], lty=2)

axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))

```

(最新3年分予測)平滑化トレンドモデル+変動季節変動+前年同期比



- モデル6（平滑化トレンドモデル+可変季節変動+前年同期比（対数変換あり））では、各月の不定期な変動要素にも対応しつつ予測していることが伺える一方、モデル4（平滑化トレンドモデル+可変季節変動（対数変換なし））では、一定の季節変動のもとでの予測であることが伺える。
- よって、今回トライした中では、モデルの当てはまり、予測の両面でベターなモデルは モデル6：平滑化トレンドモデル+可変季節変動+前年同期比（対数変換あり）である。