

Empirical Finance: Tutorial - Week 6

Tomás Cordeiro da Silva

2024-02-12

Contents

Setup	1
Working Directory	1
Housekeeping	2
Packages	2
Functions	2
Exercise 1	2
Data	2
OLS	4
DOLS	4
OLS vs DOLS	5
Exercise 2	6
Data	6
OOS Economic Evaluation	7
Model	8
Benchmark	8
Evaluation	8
Economic Restrictions	11
Sign Restriction 1	11
Sign Restriction 2	13
Sign Restriction 3	15
Evaluation	17

Setup

Working Directory

```
# --- confirm working directory
getwd()

## [1] "/Users/rodologie/Documents/Imperial/Term2/Empirical Finance/TA/TA6/PS5_Solutions"

# --- list existing files
list.files()

## [1] "data"          "functions"      "tutorial_week6_files"
## [4] "tutorial_week6.html" "tutorial_week6.Rmd" "tutorial_week6.Rproj"
```

Housekeeping

```
# --- removes all objects from the the current environment
rm(list = ls())
```

Packages

```
# --- defining the packages required
packages <- c(
  "tidyverse", # data manipulation, transformation, and visualization
  "readxl",    # import of Excel files (.xls, .xlsx)
  "sandwich",
  "knitr"
)

# --- from the ones needed extract the ones not installed
to_install <- packages[!packages %in% installed.packages()[, "Package"]]

# --- install the packages
if (length(to_install) > 0) {
  install.packages(to_install)
}

# --- load the packages in our session
invisible(lapply(packages, library, character.only = TRUE))
```

Functions

```
source("functions/log_excess_return.R")
```

Exercise 1

Data

```
cay <- read.csv("data/cay.csv")

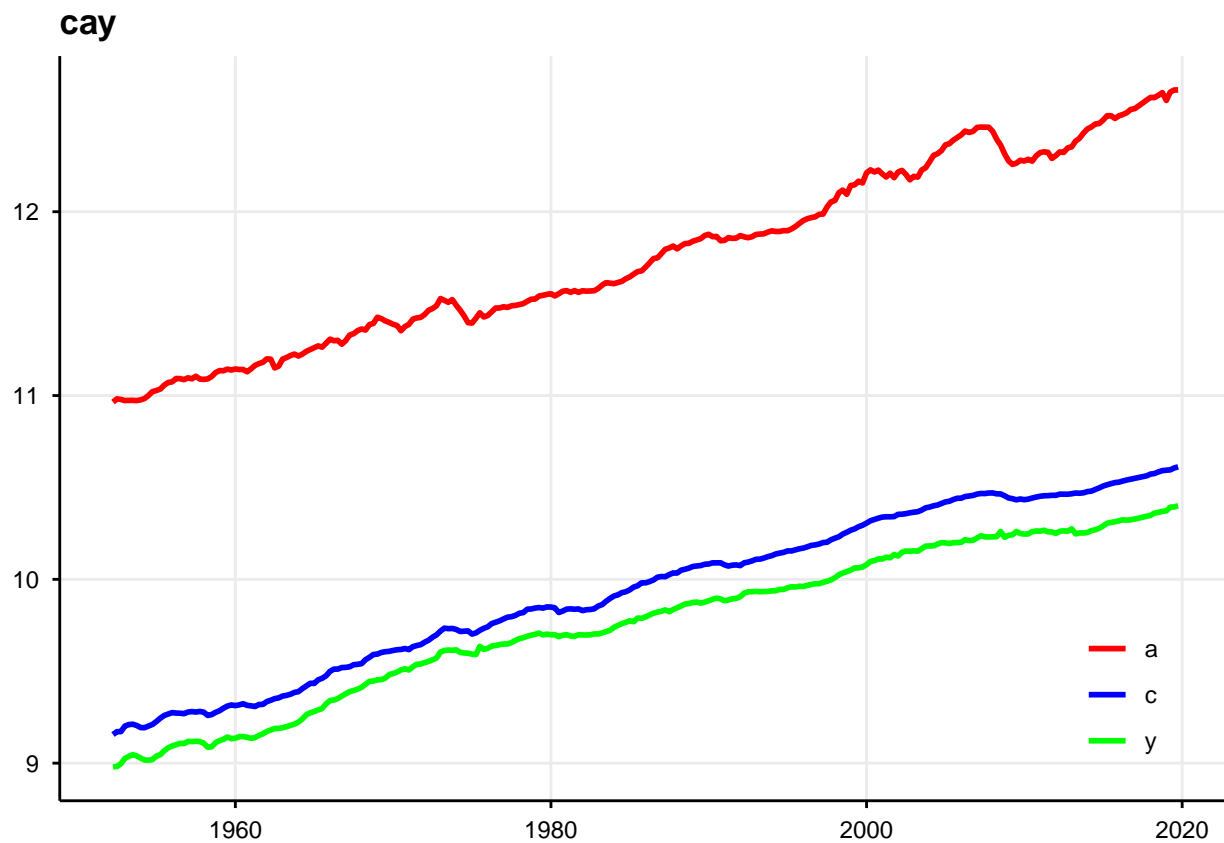
cay <- cay %>%
  mutate(
    date = as.Date(dmy(date)),
    c     = log(C),
    a     = log(A),
    y     = log(Y),
    da_lead = lead(a,1) - a,
    da_lag  = a - lag(a,1),
    dy_lead = lead(y,1) - y,
    dy_lag  = y - lag(y,1)
  )

ggplot(cay, aes(x = date)) +
  geom_line(aes(y = c, color = "c"), linewidth = 1) +
  geom_line(aes(y = a, color = "a"), linewidth = 1) +
  geom_line(aes(y = y, color = "y"), linewidth = 1) +
```

```

scale_color_manual(
  values = c(
    "c" = "blue",
    "a" = "red",
    "y" = "green"
  ),
  # labels = c(
  #   "c" = "Consumption",
  #   "a" = "Assets",
  #   "y" = "Income"
  # ),
  guide = guide_legend(title = NULL)
) +
labs(
  x = NULL,
  y = NULL,
  title = "cay"
) +
theme_minimal() +
theme(
  axis.text = element_text(color = "black"),
  axis.title = element_text(color = "black"),
  axis.line = element_line(color = "black"),
  axis.ticks = element_line(color = "black"),
  plot.title = element_text(face = "bold"),
  # panel.grid = element_blank(),
  panel.grid.minor.x = element_blank(),
  panel.grid.minor.y = element_blank(),
  legend.position = c(0.95, 0.03),
  legend.justification = c(1, 0)
)

```



OLS

$$c_t = \alpha + \beta_a a_t - \beta_y y_t + \epsilon_t$$

$$cay_t = c_t - \hat{\beta}_a a_t - \hat{\beta}_y y_t$$

```
model <- lm(c ~ 1 + a + y, cay)

# --- compute the Newey-West robust standard errors
nw_se <- NeweyWest(model)

# --- extract coefficients
b <- coef(model)

# --- compute cay (residuals)
cay$cay_OLS <- cay$c - b[1] - b[2]*cay$a - b[3] * cay$y
```

DOLS

$$c_t = \alpha + \beta_a a_t + \beta_y y_t + \sum_{i=-1}^1 b_{a,i} \Delta a_{t-i} + \sum_{i=-1}^1 b_{y,i} \Delta y_{t-i} + \epsilon_t$$

$$\begin{aligned}
\sum_{i=-1}^1 b_{a,i} \Delta a_{t-i} &= b_{a,-1} \Delta a_{t+1} + b_{a,0} \Delta a_t + b_{a,1} \Delta a_{t-1} \\
&= b_{a,-1} (a_t + 1 - a_t) + b_{a,0} \underbrace{(a_t - a_t)}_{=0} + b_{a,1} (a_t - a_{t-1}) \\
&= b_{a,-1} \underbrace{(a_t + 1 - a_t)}_{\text{lead}} + b_{a,1} \underbrace{(a_t - a_{t-1})}_{\text{lag}}
\end{aligned}$$

$$c_t = \alpha + \beta_a a_t + \beta_y y_t + b_{a,-1} (a_t + 1 - a_t) + b_{a,1} (a_t - a_{t-1}) + b_{y,-1} (y_t + 1 - y_t) + b_{y,1} (y_t - y_{t-1}) + \epsilon_t$$

$$\text{cay}_t = c_t - \hat{\beta}_a a_t - \hat{\beta}_y y_t$$

```

model <- lm(c ~ 1 + a + y + da_lead + da_lag + dy_lead + dy_lag, na.omit(cay))

# --- compute the Newey-West robust standard errors
nw_se <- NeweyWest(model)

# --- extract coefficients
b <- coef(model)

# --- compute cay (residuals)
cay$cay_DOLS <- cay$c - b[1] - b[2]*cay$a - b[3] * cay$y

```

OLS vs DOLS

```

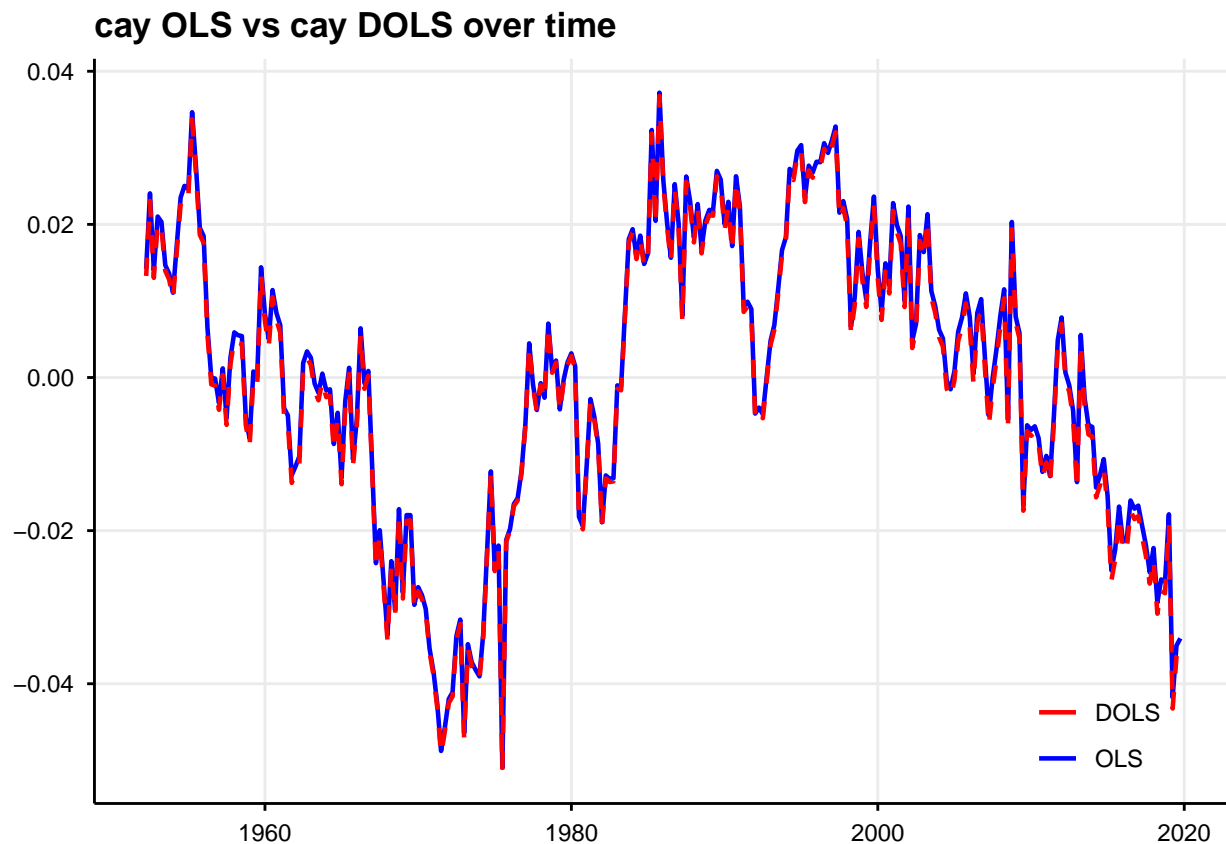
ggplot(cay, aes(x = date)) +
  geom_line(aes(y = cay_OLS, color = "cay_OLS", linetype = "cay_OLS"), linewidth = 0.8) +
  geom_line(aes(y = cay_DOLS, color = "cay_DOLS", linetype = "cay_DOLS"), linewidth = 0.8) +
  scale_color_manual(
    values = c(
      "cay_OLS" = "blue",
      "cay_DOLS" = "red"
    ),
    labels = c(
      "cay_OLS" = "OLS",
      "cay_DOLS" = "DOLS"
    ),
    guide = guide_legend(title = NULL)
  ) +
  scale_linetype_manual(
    values = c(
      "cay_OLS" = "solid",
      "cay_DOLS" = "dashed"
    ),
    guide = "none"
  ) +
  labs(
    x = NULL,
    y = NULL,
    title = "cay OLS vs cay DOLS over time"
  ) +

```

```

theme_minimal() +
theme(
  axis.text = element_text(color = "black"),
  axis.title = element_text(color = "black"),
  axis.line = element_line(color = "black"),
  axis.ticks = element_line(color = "black"),
  plot.title = element_text(face = "bold"),
  # panel.grid = element_blank(),
  panel.grid.minor.x = element_blank(),
  panel.grid.minor.y = element_blank(),
  legend.position = c(0.95, 0.01),
  legend.justification = c(1, 0)
)

```



Exercise 2

Data

```

# --- stock index
stock <- read.csv("data/stock_index.csv")

stock <- stock %>%
  mutate(
    date = as.Date(dmy(date)),
    period = paste(year(date), "-", month(date))
  ) %>%

```

```

select(-date)

# --- riskless rate
tbill <- read.csv("data/riskless_rate.csv")

tbill <- tbill %>%
  mutate(
    date = as.Date(dmy(date)),
    period = paste(year(date), "-", month(date))
  ) %>%
  select(-date)

# --- cay
cay_OLS <- cay %>%
  select(date, cay_OLS) %>%
  mutate(period = paste(year(date), "-", month(date)))

# --- merging the data
data <- cay_OLS %>%
  inner_join(
    stock,
    by = "period"
  ) %>%
  inner_join(
    tbill,
    by = "period"
  ) %>%
  select(-period)

head(data)

##           date    cay_OLS    price riskless
## 1 1952-03-31 0.01420660 24.08287      1.50
## 2 1952-06-30 0.02405367 24.65940      1.74
## 3 1952-09-30 0.01394102 25.06398      1.60
## 4 1952-12-31 0.02102993 26.33842      2.22
## 5 1953-03-31 0.02032302 26.28785      1.97
## 6 1953-06-30 0.01466728 24.22447      1.89

# --- extracting the dates vector
date <- data$date

# --- converting to matrix
stock <- matrix(data$price)
tbill <- matrix(data$riskless)
cay <- matrix(data$cay_OLS)

```

OOS Economic Evaluation

```

# --- computing log returns (keeping the first observation missing) and bringing annualized rates to mo
ex <- log_excess_return(stock, tbill/400)
rf <- tbill/400

# --- parameters

```

```
window <- 40 # window size
```

```
# --- dimensions
```

```
T <- dim(cay)[1]
```

```
N <- dim(cay)[2]
```

Model

```
# --- initialize storage objects
```

```
mod_fmat = matrix(-9999.0, T)
```

```
mod_emat = matrix(-9999.0, T)
```

```
mod_bmat = matrix(-9999.0, T)
```

```
j <- 0
```

```
for(t in window:(T-1)) {
```

```
  Y <- ex[(2+j):t]
```

```
  X <- cbind(1, cay[(1+j):(t-1)])
```

```
  b <- solve(t(X) %*% X) %*% (t(X) %*% Y)
```

```
  mod_bmat[t+1] <- b[2]
```

```
  mod_fmat[t+1] <- b[1] + b[2]*cay[t]
```

```
  mod_emat[t+1] <- ex[t+1] - (b[1] + b[2]*cay[t])
```

```
  j <- j + 1
```

```
}
```

Benchmark

```
# --- initialize storage objects
```

```
ben_fmat = matrix(-9999.0, T)
```

```
ben_emat = matrix(-9999.0, T)
```

```
j <- 0
```

```
for(t in window:(T-1)) {
```

```
  Y <- ex[(2+j):t]
```

```
  X <- cbind(1, cay[(1+j):(t-1)])
```

```
  b <- mean(Y)
```

```
  ben_fmat[t+1] <- b
```

```
  ben_emat[t+1] <- ex[t+1] - b
```

```
  j <- j + 1
```

```
}
```

Evaluation

MSE


```
# ---
ind <- which(mod_fmat != -9999.0)

# --- Competing model MSE
MSE_mod <- mean(mod_emat[ind]^2, na.rm = TRUE)

# --- Benchmark MSE
MSE_ben <- mean(ben_emat[ind]^2, na.rm = TRUE)
```

OOS R2

```
# --- OOS R_squared
OOS_R2 <- 1 - (MSE_mod/MSE_ben)
```

Clark-West Test

```
# --- step 1
fstat <- (ben_emat[ind]^2) - ((mod_emat[ind]^2) - (ben_fmat[ind] - mod_fmat[ind])^2)

# --- step 2
model <- lm(fstat ~ 1)

# --- step 3
rho_hat <- coef(model)[[1]]
se <- summary(model)$coefficients[1, 2]

CW <- rho_hat / se
```

Summary

```
summary <- data.frame(
  Metric = c(
    "MSE_mod",
    "MSE_ben",
    "OOS_R2",
    "CW"
  ),
  `Rolling Window` = c(
    MSE_mod,
    MSE_ben,
    OOS_R2,
    CW
  ),
  check.names = FALSE
)

latex_labels <- c(
  MSE_mod = "$MSE_{MOD}$",
  MSE_ben = "$MSE_{BEN}$",
  OOS_R2 = "$OOS\\ R^2$",
  CW = "$CW$"
)

summary$Metric <- latex_labels[summary$Metric]

kable(
```

```
summary,
caption = "Out-of-Sample Forecast Performance",
digits = 5
)
```

Table 1: Out-of-Sample Forecast Performance

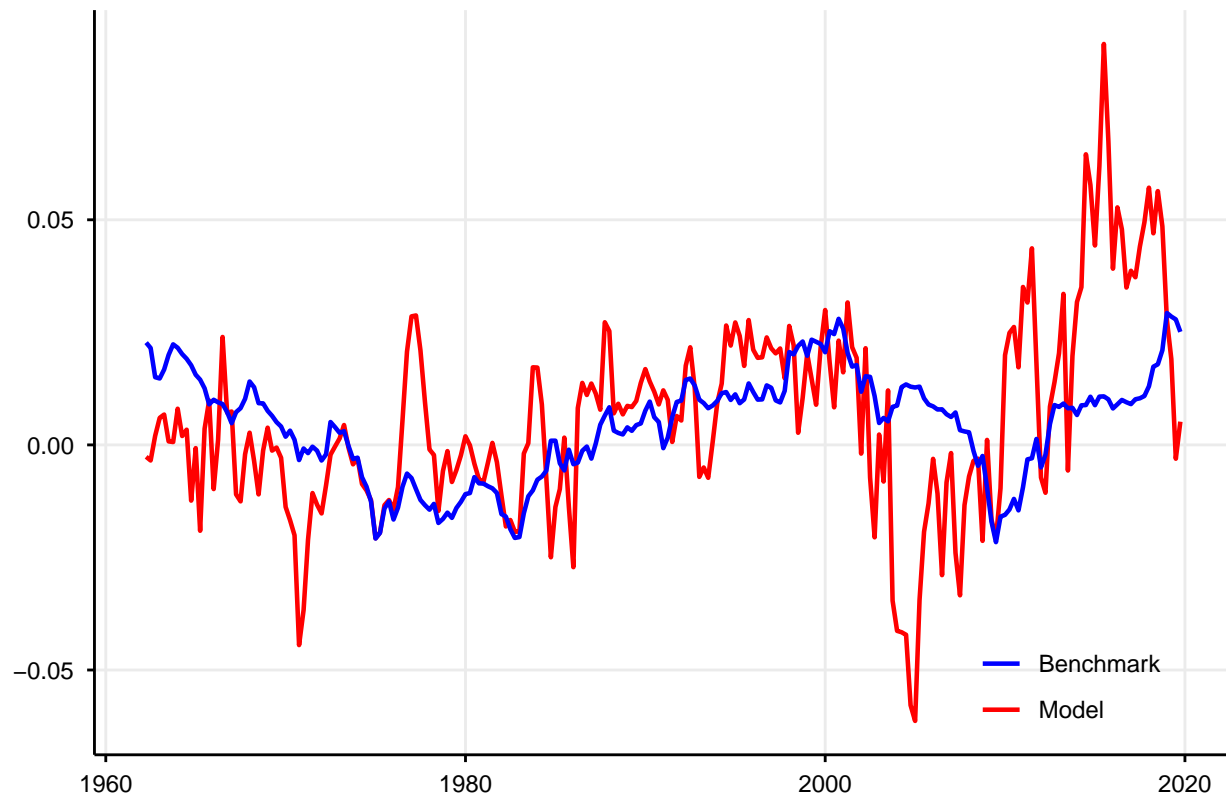
Metric	Rolling Window
MSE_{MOD}	0.00688
MSE_{BEN}	0.00661
$OOS R^2$	-0.04079
CW	1.08628

```
fmat <- data.frame(
  date      = cay_OLS$date,
  observed  = ex,
  mod_f     = mod_fmat,
  ben_f     = ben_fmat
) %>%
filter(mod_f != -9999.0)

ggplot(fmat, aes(x = date)) +
  # geom_line(aes(y = observed, color = "observed"), linewidth = 0.8) +
  geom_line(aes(y = mod_f,   color = "mod_f"), linewidth = 0.8) +
  geom_line(aes(y = ben_f,   color = "ben_f"), linewidth = 0.8) +
  scale_color_manual(
    values = c(
      # observed = "black",
      mod_f     = "red",
      ben_f     = "blue"
    ),
    labels = c(
      # observed = "y",
      mod_f     = "Model",
      ben_f     = "Benchmark"
    ),
    guide = guide_legend(title = NULL)
) +
labs(
  x = NULL,
  y = NULL,
  title = "Forecast Comparison"
) +
theme_minimal() +
theme(
  axis.text = element_text(color = "black"),
  axis.title = element_text(color = "black"),
  axis.line = element_line(color = "black"),
  axis.ticks = element_line(color = "black"),
  plot.title = element_text(face = "bold"),
  panel.grid.minor = element_blank(),
  legend.position = c(0.95, 0.01),
```

```
legend.justification = c(1, 0)
)
```

Forecast Comparison



Economic Restrictions

Sign Restriction 1

Set $\hat{\beta} = 0$ when $\hat{\beta} < 0$

```
# --- initialize storage objects
mod_fmat_1 = matrix(-9999.0, T)
mod_emat_1 = matrix(-9999.0, T)
mod_bmat_1 = matrix(-9999.0, T)

j <- 0

for(t in window:(T-1)) {

  Y <- ex[(2+j):t]
  X <- cbind(1, cay[(1+j):(t-1)])

  b <- solve(t(X) %*% X) %*% (t(X) %*% Y)

  # -- impose sign restriction 1
  if (b[2] < 0){b[2] = 0}

  mod_bmat_1[t+1] <- b[2]
```

```

mod_fmat_1[t+1] <- b[1] + b[2]*cay[t]
mod_emat_1[t+1] <- ex[t+1] - (b[1] + b[2]*cay[t])

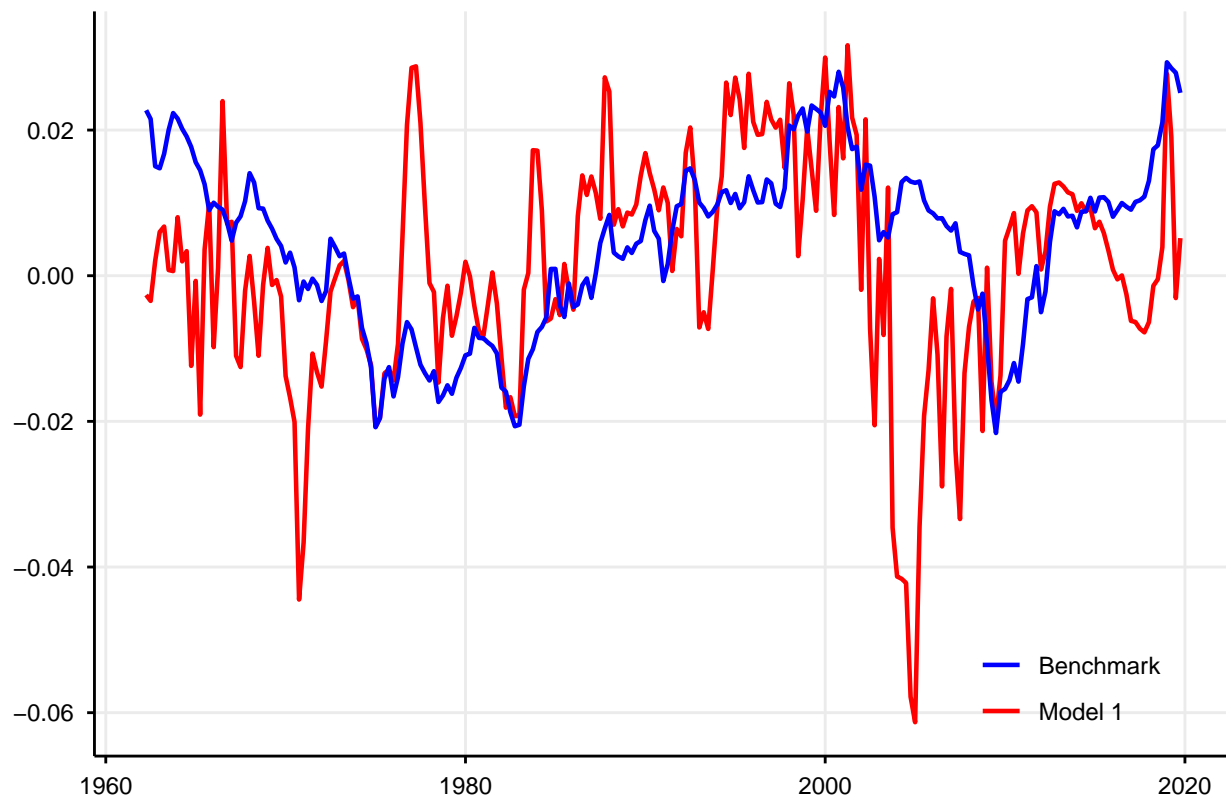
j <- j + 1
}

fmat <- data.frame(
  date      = cay_OLS$date,
  observed  = ex,
  mod_f     = mod_fmat_1,
  ben_f     = ben_fmat
) %>%
filter(mod_f != -9999.0)

ggplot(fmat, aes(x = date)) +
  # geom_line(aes(y = observed, color = "observed"), linewidth = 0.8) +
  geom_line(aes(y = mod_f, color = "mod_f"), linewidth = 0.8) +
  geom_line(aes(y = ben_f, color = "ben_f"), linewidth = 0.8) +
  scale_color_manual(
    values = c(
      # observed = "black",
      mod_f      = "red",
      ben_f      = "blue"
    ),
    labels = c(
      # observed = "y",
      mod_f      = "Model 1",
      ben_f      = "Benchmark"
    ),
    guide = guide_legend(title = NULL)
  ) +
  labs(
    x = NULL,
    y = NULL,
    title = "Forecast Comparison"
  ) +
  theme_minimal() +
  theme(
    axis.text = element_text(color = "black"),
    axis.title = element_text(color = "black"),
    axis.line = element_line(color = "black"),
    axis.ticks = element_line(color = "black"),
    plot.title = element_text(face = "bold"),
    panel.grid.minor = element_blank(),
    legend.position = c(0.95, 0.01),
    legend.justification = c(1, 0)
  )

```

Forecast Comparison



Sign Restriction 2

Set $\hat{y}_{t+1} = 0$ when $\hat{y}_{t+1} < 0$

```
# --- initialize storage objects
mod_fmat_2 = matrix(-9999.0, T)
mod_emat_2 = matrix(-9999.0, T)
mod_bmat_2 = matrix(-9999.0, T)

j <- 0

for(t in window:(T-1)) {

  Y <- ex[(2+j):t]
  X <- cbind(1, cay[(1+j):(t-1)])

  b <- solve(t(X) %*% X) %*% (t(X) %*% Y)

  f <- b[1] + b[2]*cay[t]

  # -- impose sign restriction 2
  if (f < 0){f = 0}

  mod_bmat_2[t+1] <- b[2]
  mod_fmat_2[t+1] <- f
  mod_emat_2[t+1] <- ex[t+1] - f
}
```

```

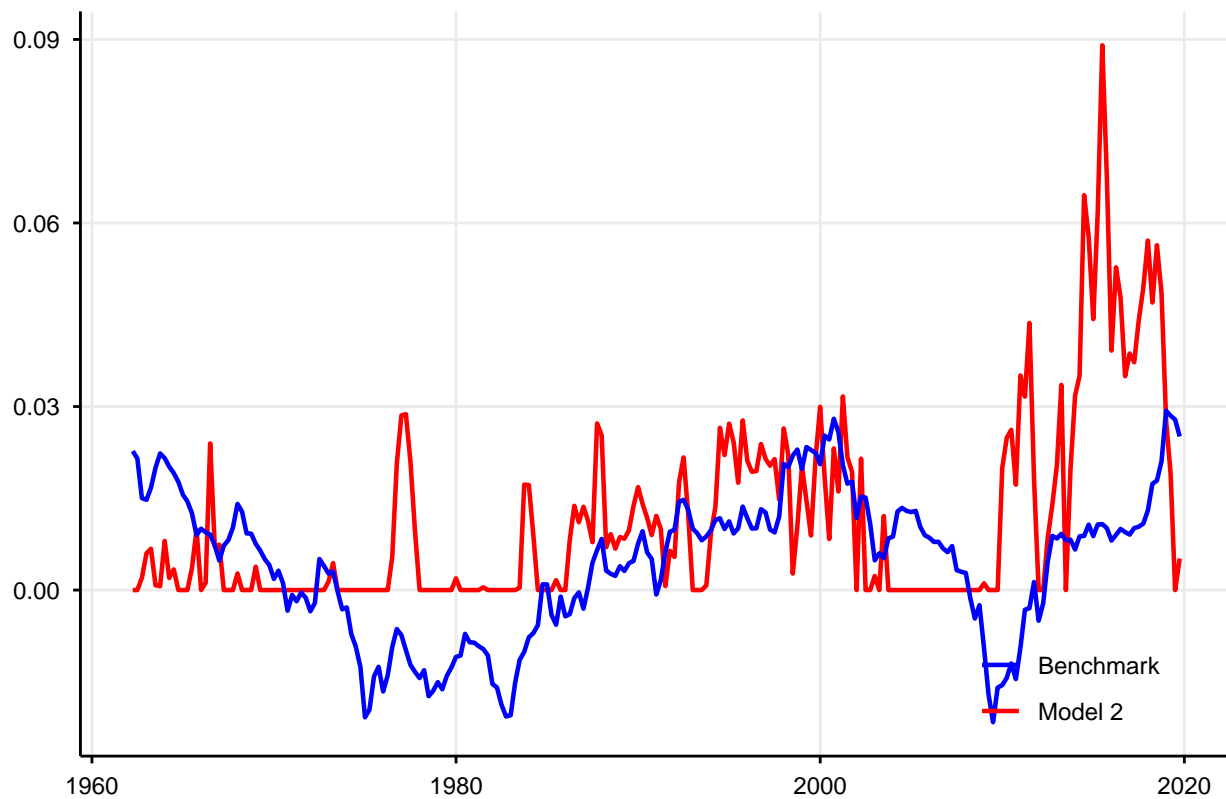
    j <- j + 1
  }

fmat <- data.frame(
  date      = cay_OLS$date,
  observed  = ex,
  mod_f     = mod_fmat_2,
  ben_f     = ben_fmat
) %>%
filter(mod_f != -9999.0)

ggplot(fmat, aes(x = date)) +
  # geom_line(aes(y = observed, color = "observed"), linewidth = 0.8) +
  geom_line(aes(y = mod_f,    color = "mod_f"), linewidth = 0.8) +
  geom_line(aes(y = ben_f,    color = "ben_f"), linewidth = 0.8) +
  scale_color_manual(
    values = c(
      # observed = "black",
      mod_f      = "red",
      ben_f      = "blue"
    ),
    labels = c(
      # observed = "y",
      mod_f      = "Model 2",
      ben_f      = "Benchmark"
    ),
    guide = guide_legend(title = NULL)
  ) +
  labs(
    x = NULL,
    y = NULL,
    title = "Forecast Comparison"
  ) +
  theme_minimal() +
  theme(
    axis.text = element_text(color = "black"),
    axis.title = element_text(color = "black"),
    axis.line = element_line(color = "black"),
    axis.ticks = element_line(color = "black"),
    plot.title = element_text(face = "bold"),
    panel.grid.minor = element_blank(),
    legend.position = c(0.95, 0.01),
    legend.justification = c(1, 0)
  )

```

Forecast Comparison



Sign Restriction 3

Set $\hat{\beta} = 0$ when $\hat{\beta} < 0$ and $\hat{y}_{t+1} = 0$ when $\hat{y}_{t+1} < 0$

```
# --- initialize storage objects
mod_fmat_3 = matrix(-9999.0, T)
mod_emat_3 = matrix(-9999.0, T)
mod_bmat_3 = matrix(-9999.0, T)

j <- 0

for(t in window:(T-1)) {

  Y <- ex[(2+j):t]
  X <- cbind(1, cay[(1+j):(t-1)])

  b <- solve(t(X) %*% X) %*% (t(X) %*% Y)

  # -- impose sign restriction 3
  if (b[2] < 0){b[2] = 0}

  f <- b[1] + b[2]*cay[t]

  # -- impose sign restriction 3
  if (f < 0){f = 0}

  mod_bmat_3[t+1] <- b[2]
```

```

mod_fmat_3[t+1] <- f
mod_emat_3[t+1] <- ex[t+1] - f

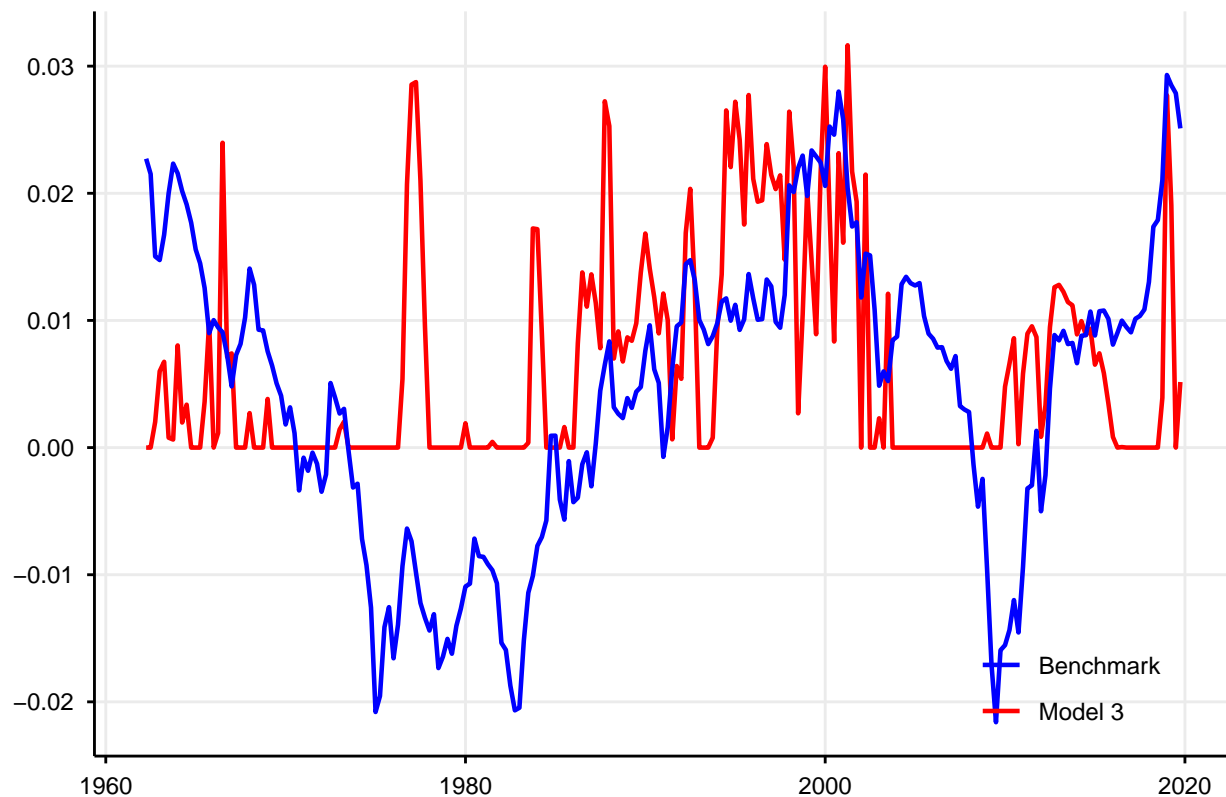
j <- j + 1
}

fmat <- data.frame(
  date      = cay_OLS$date,
  observed  = ex,
  mod_f     = mod_fmat_3,
  ben_f     = ben_fmat
) %>%
filter(mod_f != -9999.0)

ggplot(fmat, aes(x = date)) +
  # geom_line(aes(y = observed, color = "observed"), linewidth = 0.8) +
  geom_line(aes(y = mod_f, color = "mod_f"), linewidth = 0.8) +
  geom_line(aes(y = ben_f, color = "ben_f"), linewidth = 0.8) +
  scale_color_manual(
    values = c(
      # observed = "black",
      mod_f     = "red",
      ben_f     = "blue"
    ),
    labels = c(
      # observed = "y",
      mod_f     = "Model 3",
      ben_f     = "Benchmark"
    ),
    guide = guide_legend(title = NULL)
  ) +
  labs(
    x = NULL,
    y = NULL,
    title = "Forecast Comparison"
  ) +
  theme_minimal() +
  theme(
    axis.text = element_text(color = "black"),
    axis.title = element_text(color = "black"),
    axis.line = element_line(color = "black"),
    axis.ticks = element_line(color = "black"),
    plot.title = element_text(face = "bold"),
    panel.grid.minor = element_blank(),
    legend.position = c(0.95, 0.01),
    legend.justification = c(1, 0)
  )

```


Forecast Comparison



Evaluation

MSE

```
MSE_mod_1 <- mean(mod_emat_1[ind]^2, na.rm = TRUE)
MSE_mod_2 <- mean(mod_emat_2[ind]^2, na.rm = TRUE)
MSE_mod_3 <- mean(mod_emat_3[ind]^2, na.rm = TRUE)
```

OOS R2

```
OOS_R2_1 <- 1 - (MSE_mod_1/MSE_ben)
OOS_R2_2 <- 1 - (MSE_mod_2/MSE_ben)
OOS_R2_3 <- 1 - (MSE_mod_3/MSE_ben)
```

Clark-West Test

```
# --- step 1
fstat_1 <- (ben_emat[ind]^2) - ((mod_emat_1[ind]^2) - (ben_fmat[ind] - mod_fmat_1[ind])^2)
fstat_2 <- (ben_emat[ind]^2) - ((mod_emat_2[ind]^2) - (ben_fmat[ind] - mod_fmat_2[ind])^2)
fstat_3 <- (ben_emat[ind]^2) - ((mod_emat_3[ind]^2) - (ben_fmat[ind] - mod_fmat_3[ind])^2)

# --- step 2
model_1 <- lm(fstat_1 ~ 1)
model_2 <- lm(fstat_2 ~ 1)
model_3 <- lm(fstat_3 ~ 1)

# --- step 3
rho_hat_1 <- coef(model_1)[[1]]
se_1 <- summary(model_1)$coefficients[1, 2]
```

```

CW_1 <- rho_hat_1 / se_1

rho_hat_2 <- coef(model_2)[[1]]
se_2      <- summary(model_2)$coefficients[1, 2]

CW_2 <- rho_hat_2 / se_2

rho_hat_3 <- coef(model_3)[[1]]
se_3      <- summary(model_3)$coefficients[1, 2]

CW_3 <- rho_hat_3 / se_3

```

Summary

```

summary <- data.frame(
  Metric = c(
    "MSE_mod",
    "MSE_ben",
    "OOS_R2",
    "CW"
  ),
  `No Sign Restrictions` = c(
    MSE_mod,
    MSE_ben,
    OOS_R2,
    CW
  ),
  `Sign Restriction 1` = c(
    MSE_mod_1,
    MSE_ben,
    OOS_R2_1,
    CW_1
  ),
  `Sign Restriction 2` = c(
    MSE_mod_2,
    MSE_ben,
    OOS_R2_2,
    CW_2
  ),
  `Sign Restriction 3` = c(
    MSE_mod_3,
    MSE_ben,
    OOS_R2_3,
    CW_3
  ),
  check.names = FALSE
)

latex_labels <- c(
  MSE_mod = "$MSE_{MOD}$",
  MSE_ben = "$MSE_{BEN}$",
  OOS_R2  = "$OOS\\ R^2$",
  CW      = "$CW$"
)

```

```
summary$Metric <- latex_labels[summary$Metric]

kable(
  summary,
  caption = "Out-of-Sample Forecast Performance",
  digits = 5
)
```

Table 2: Out-of-Sample Forecast Performance

Metric	No Sign Restrictions	Sign Restriction 1	Sign Restriction 2	Sign Restriction 3
MSE_{MOD}	0.00688	0.00679	0.00657	0.00652
MSE_{BEN}	0.00661	0.00661	0.00661	0.00661
$OOS R^2$	-0.04079	-0.02814	0.00552	0.01272
CW	1.08628	0.63829	2.24938	1.69890