HW1 Associative Model Forecast

Group 2

2025-09-13

Table of contents

中文

1 Explain Data

已對資料做整理，取每個月的最後一筆收盤價當成月的基礎。

再利用轉對數的形式讓小樣本的序列能更平穩

報酬率公式:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

利用報酬率來預測輝達收盤價跟SOXX(美國半導體產業股票指數)的收盤價的關聯

本次使用ARIMAX 模型來探討，但因資料筆數為11筆，模型只能作為參考。

```r
library(readr)
library(forecast)
```

```
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```

```r
library(skimr)
library(knitr)
library(kableExtra)
data <- read_csv("NVDA_SOXX_monthly_prices_and_logreturns.csv")
```

```
Rows: 11 Columns: 5

-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): MonthEnd
dbl (4): NVDA, SOXX, NVDA_ret, SOXX_ret

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
skim(data)
```

Table 1: Data summary

Name	data
Number of rows	11
Number of columns	5
Column type frequency:	
character	1
numeric	4
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
MonthEnd	0	1	9	10	0	11	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
NVDA	0	1	137.52	23.58	108.38	122.50	134.29	148.12	177.87	
SOXX	0	1	216.07	19.87	183.84	206.73	215.49	228.48	245.32	
NVDA_ret	0	1	0.03	0.11	-0.14	-0.03	0.04	0.10	0.22	
SOXX_ret	0	1	0.01	0.07	-0.10	-0.03	0.00	0.02	0.15	

2 Test

下面使用adf test 來檢驗，因資料筆數的不夠，

導致p-value值過高或出現NA

```
library(tseries)
library(knitr)
library(broom)
nvda_adf <- adf.test(data$NVDA_ret)
soxx_adf <- adf.test(data$SOXX_ret)
kable(tidy(nvda_adf), caption = "ADF test on NVDA returns", digits = 4)
```

Table 4: ADF test on NVDA returns

statistic	p.value	parameter	method	alternative
-1.8792	0.6184	2	Augmented Dickey-Fuller Test	stationary

```
kable(tidy(soxx_adf), caption = "ADF test on SOXX returns", digits = 4)
```

Table 5: ADF test on SOXX returns

statistic	p.value	parameter	method	alternative
-0.8332	0.9451	2	Augmented Dickey-Fuller Test	stationary

3 ARIMAX model

下面讓我們的輝達股價為 Y ，讓SOXX的股價無X ，

由於ARIMA(0,0,0)，所以在ARIMAX裡只做簡單線性回歸，因資料筆數太少

探討是否因為SOXX的股票升漲，進而影響輝達的股價，

可以看出在 ARIMAX 模型中，SOXX 的迴歸係數估計為 1.0253（標準誤 0.3631，t 值 □ 2.82，顯著），

這表示在控制自相關結構後，SOXX 報酬與 NVDA 報酬之間存在顯著的正向關聯：

SOXX 每上漲 1%，NVDA 約上漲 1.025%。

這符合金融直覺，因為 SOXX 代表整體半導體產業，而 NVDA 作為產業龍頭

，其股價自然與產業走勢緊密連動。

誤差項可以看出ACF1：殘差一階自相關 □ -0.19，代表殘差大致沒有強烈自相關。

百分比誤差超過150%應該是因為 y_ts 值太接近零（導致分母小）。

```
y <- data$NVDA_ret
X <- stats::lag(data$SOXX_ret, -1)   #    1   t    NVDA    t-1    SOXX
df <- data.frame(y=y, X=X)
df <- na.omit(df)

y_ts <- ts(df$y, frequency = 12, start = c(2024,8))  #   2024   8
X_mat <- as.matrix(df$X)
# ----    ARIMAX ----
fit <- auto.arima(y_ts, xreg = X_mat, seasonal = FALSE,
                  stepwise = FALSE, approximation = FALSE)
library(broom)
library(knitr)
library(kableExtra)

#
coef_tab <- tidy(fit)
kable(coef_tab, format="latex", booktabs=TRUE,
      caption="ARIMAX Model Coefficients") %>%
  kable_styling(latex_options="scale_down", position="center")
```

```
#    (AIC, BIC, logLik)
model_info <- glance(fit)
kable(model_info, format="latex", booktabs=TRUE,
      caption="ARIMAX Model Fit Summary") %>%
  kable_styling(latex_options="scale_down", position="center")
```

Table 6: ARIMAX Model Coefficients

term	estimate	std.error
xreg	1.025308	0.3630819

Table 7: ARIMAX Model Fit Summary

sigma	logLik	AIC	BIC	nobs
0.0868561	11.7944	-19.58881	-18.79302	11

```
bp <- tidy(Box.test(fit$residuals, type="Box-Pierce", lag=1))
kable(bp, format="latex", booktabs=TRUE,
      caption="Box-Pierce Test on Residuals") %>%
  kable_styling(latex_options="scale_down", position="center")
```

4 Forecast

下面用報酬預測9~12月的預測再轉為收盤價得出結果，使用了天真法預測。

```
future_X <- matrix(rep(tail(df$X,1), 4), ncol = 1)
colnames(future_X) <- "X"
fc <- forecast(fit, xreg = future_X, h=4)
```

```
Warning in forecast.forecast_ARIMA(fit, xreg = future_X, h = 4): xreg contains
different column names from the xreg used in training. Please check that the
regressors are in the same order.
```

```
library(knitr)
library(kableExtra)

fc_table <- data.frame(
  Step   = 1:length(fc$mean),
```

Table 8: Box-Pierce Test on Residuals

statistic	p.value	parameter	method
0.3991365	0.5275356	1	Box-Pierce test

Table 9: NVDA Forecast (h = 4)

Step	Mean	Lo80	Hi80	Lo95	Hi95
1	0.022351	-0.088959	0.133662	-0.147884	0.192586
2	0.022351	-0.088959	0.133662	-0.147884	0.192586
3	0.022351	-0.088959	0.133662	-0.147884	0.192586
4	0.022351	-0.088959	0.133662	-0.147884	0.192586

```
  Mean  = round(as.numeric(fc$mean), 6),
  Lo80  = round(as.numeric(fc$lower[, "80%"]), 6),
  Hi80  = round(as.numeric(fc$upper[, "80%"]), 6),
  Lo95  = round(as.numeric(fc$lower[, "95%"]), 6),
  Hi95  = round(as.numeric(fc$upper[, "95%"]), 6)
)

kable(
  fc_table,
  format = "latex", booktabs = TRUE,
  caption = "NVDA Forecast (h = 4)",
  align = c("c", rep("r", 5))
) %>%
  kable_styling(latex_options = "scale_down", position = "center")
```

```
h <- 4
ret4 <- as.numeric(tail(fc$mean, h))

#   →
last_price <- as.numeric(tail(data$NVDA, 1))
price4 <- last_price * cumprod(exp(ret4))

#    =4
pred_table <- data.frame(
  Month = c("2025-09", "2025-10", "2025-11", "2025-12"),
  NVDA_ret_forecast   = ret4,
  NVDA_price_forecast = price4,
  stringsAsFactors = FALSE
)



library(knitr)
library(kableExtra)
```

Month	NVDA_ret_forecast	NVDA_price_forecast
2025-09	0.0223513	178.1170
2025-10	0.0223513	182.1429
2025-11	0.0223513	186.2599
2025-12	0.0223513	190.4699

```r
library(dplyr)
```

```
Attaching package: 'dplyr'


The following object is masked from 'package:kableExtra':

    group_rows


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
kable(
  pred_table %>%
    mutate(
      NVDA_ret_forecast   = round(NVDA_ret_forecast, 8),
      NVDA_price_forecast = round(NVDA_price_forecast, 4)
    ),
  format = "latex", booktabs = TRUE,
  col.names = c("Month", "NVDA_ret_forecast", "NVDA_price_forecast"),
  align = c("c","r","r")
) %>%
  kable_styling(latex_options = "scale_down", position = "center")
```

5 Forecast 2

這段使用了蒙地卡羅的方法預測，展開了不確定性，模擬多條SOXX的可能路徑 ，處理了未來外生變數
在傳遞到NVDA的ARIMAX。

我們先對SOXX進行ARIMA預估，再丟入NVDA的ARIMAX進行股價的預測。

並產生三個悲觀和正常和樂觀的三個部分

```
set.seed(42)    #
h <- 4          #    NVDA 4
#   NVDA_t   SOXX_{t-1}     3    10~12


y <- data$NVDA_ret
X <- stats::lag(data$SOXX_ret, -1)        #   t   NVDA   t-1   SOXX
df <- na.omit(data.frame(y = y, X = X))

y_ts  <- ts(df$y,  frequency = 12, start = c(2024, 8))
X_mat <- as.matrix(df$X)

fit_nvda <- auto.arima(
  y_ts, xreg = X_mat, seasonal = FALSE,
  stepwise = FALSE, approximation = FALSE
)


# (1)   SOXX_ret   ARIMA   SOXX

soxx_ts <- ts(na.omit(data$SOXX_ret), frequency = 12, start = c(2024, 8))
fit_soxx <- auto.arima(
  soxx_ts, seasonal = FALSE,
  stepwise = FALSE, approximation = FALSE
)


# (2)      SOXX_ret   3
# simulate 2000 times

n_sim <- 2000
lastX <- as.numeric(tail(soxx_ts, 1))

#    3     SOXX ARIMA    3
#      SOXX    3
```

```r
sim_one_soxx <- function() {
  as.numeric(simulate(fit_soxx, nsim = 3, future = TRUE))
}

soxx_future_mat <- t(replicate(n_sim, sim_one_soxx()))  # n_sim x 3


# (3)       NVDA   ARIMAX
#
#         N()
#
## 95% = mean ± z * sd  →  sd   (upper-mean)/z
draw_one_from_fc <- function(fc_obj) {
  m  <- as.numeric(fc_obj$mean)
  lo <- as.numeric(fc_obj$lower[, "95%"])
  hi <- as.numeric(fc_obj$upper[, "95%"])
  z  <- qnorm(0.95)
  sd <- (hi - m) / z
  rnorm(length(m), mean = m, sd = sd)
}

nvda_ret_sims <- matrix(NA_real_, nrow = n_sim, ncol = h)

for (i in 1:n_sim) {
  # lag    1  NVDA   lastX       3
  future_X <- c(lastX, soxx_future_mat[i, ])
  fc_i <- forecast(fit_nvda, xreg = matrix(future_X, ncol = 1), h = h)
  #             ARIMAX
  nvda_ret_sims[i, ] <- draw_one_from_fc(fc_i)
}


# (4)   →

last_price <- as.numeric(tail(data$NVDA, 1))  #
#
nvda_price_sims <- t(apply(nvda_ret_sims, 1, function(r) last_price * exp(cumsum(r))))


# (5)     P10 P90

quant <- function(x, p) as.numeric(quantile(x, probs = p, na.rm = TRUE))
months <- c("2025-09","2025-10","2025-11","2025-12")
library(knitr)
```

Table 10: NVDA Forecast (Monte Carlo Simulation, 2000 runs)

Month	Ret_Mean	Ret_Median	Ret_P10	Ret_P90	Price_Mean	Price_Median	Price_P10	Price_P90
2025-09	0.022881	0.025947	-0.113550	0.157585	179.20	178.76	155.48	203.91
2025-10	0.048667	0.051294	-0.106559	0.195001	189.42	186.88	153.20	228.09
2025-11	-0.008766	-0.006962	-0.171210	0.149056	189.69	184.97	142.65	242.42
2025-12	-0.006890	-0.007120	-0.171367	0.164317	190.42	184.23	134.30	257.02

```r
library(kableExtra)

summary_table <- data.frame(
  Month = months,
  Ret_Mean = apply(nvda_ret_sims, 2, mean),
  Ret_Median = apply(nvda_ret_sims, 2, median),
  Ret_P10 = apply(nvda_ret_sims, 2, quant, p = 0.10),
  Ret_P90 = apply(nvda_ret_sims, 2, quant, p = 0.90),
  Price_Mean = apply(nvda_price_sims, 2, mean),
  Price_Median = apply(nvda_price_sims, 2, median),
  Price_P10 = apply(nvda_price_sims, 2, quant, p = 0.10),
  Price_P90 = apply(nvda_price_sims, 2, quant, p = 0.90)
)

kable(
  summary_table %>%
    mutate(
      Ret_Mean     = round(Ret_Mean, 6),
      Ret_Median   = round(Ret_Median, 6),
      Ret_P10      = round(Ret_P10, 6),
      Ret_P90      = round(Ret_P90, 6),
      Price_Mean   = round(Price_Mean, 2),
      Price_Median = round(Price_Median, 2),
      Price_P10    = round(Price_P10, 2),
      Price_P90    = round(Price_P90, 2)
    ),
  format = "latex", booktabs = TRUE,
  caption = "NVDA Forecast (Monte Carlo Simulation, 2000 runs)",
  align = c("c", rep("r", 8))
) %>%
  kable_styling(latex_options = "scale_down", position = "center")
```

```r
# -------------------------------
# 6)      bear/base/bull
#      SOXX_ret   ARIMA    ±
# -------------------------------
```

```r
#   SOXX   3     95%    →    sd
fc_soxx <- forecast(fit_soxx, h = 3)
m_sx <- as.numeric(fc_soxx$mean)
sd_sx <- (as.numeric(fc_soxx$upper[, "95%"]) - m_sx) / qnorm(0.95)

scenarios <- list(
  bear = m_sx - sd_sx,
  base = m_sx,
  bull = m_sx + sd_sx
)

run_scn <- function(path3) {
  future_X <- c(lastX, path3)
  fc <- forecast(fit_nvda, xreg = matrix(future_X, ncol = 1), h = h)
  #
  r_mean <- as.numeric(fc$mean)
  price_path <- last_price * exp(cumsum(r_mean))
  data.frame(
    Month = months,
    NVDA_ret = r_mean,
    NVDA_price = price_path
  )
}

scn_bear <- run_scn(scenarios$bear)
scn_base <- run_scn(scenarios$base)
scn_bull <- run_scn(scenarios$bull)

print_scn <- function(df, title){
  kable(
    df %>%
      mutate(
        NVDA_ret   = round(NVDA_ret, 6),
        NVDA_price = round(NVDA_price, 2)
      ),
    format = "latex", booktabs = TRUE,
    caption = paste("Scenario:", title),
    col.names = c("Month", "NVDA_ret_forecast", "NVDA_price_forecast"),
    align = c("c","r","r")
  ) %>%
    kable_styling(latex_options = "scale_down", position = "center")
}

print_scn(scn_bear, "Bear")
```

Table 11: Scenario: Bear

Month	NVDA_ret_forecast	NVDA_price_forecast
2025-09	0.022351	178.12
2025-10	-0.016030	175.28
2025-11	-0.087983	160.52
2025-12	-0.087983	147.00

Table 12: Scenario: Base

Month	NVDA_ret_forecast	NVDA_price_forecast
2025-09	0.022351	178.12
2025-10	0.053336	187.88
2025-11	0.000000	187.88
2025-12	0.000000	187.88

```r
print_scn(scn_base, "Base")
```

```r
print_scn(scn_bull, "Bull")
```

6 Plot

```r
library(dplyr)
library(ggplot2)
library(scales)
```

```
Attaching package: 'scales'
```

Table 13: Scenario: Bull

Month	NVDA_ret_forecast	NVDA_price_forecast
2025-09	0.022351	178.12
2025-10	0.122702	201.37
2025-11	0.087983	219.89
2025-12	0.087983	240.11

```
The following object is masked from 'package:readr':

    col_factor
```

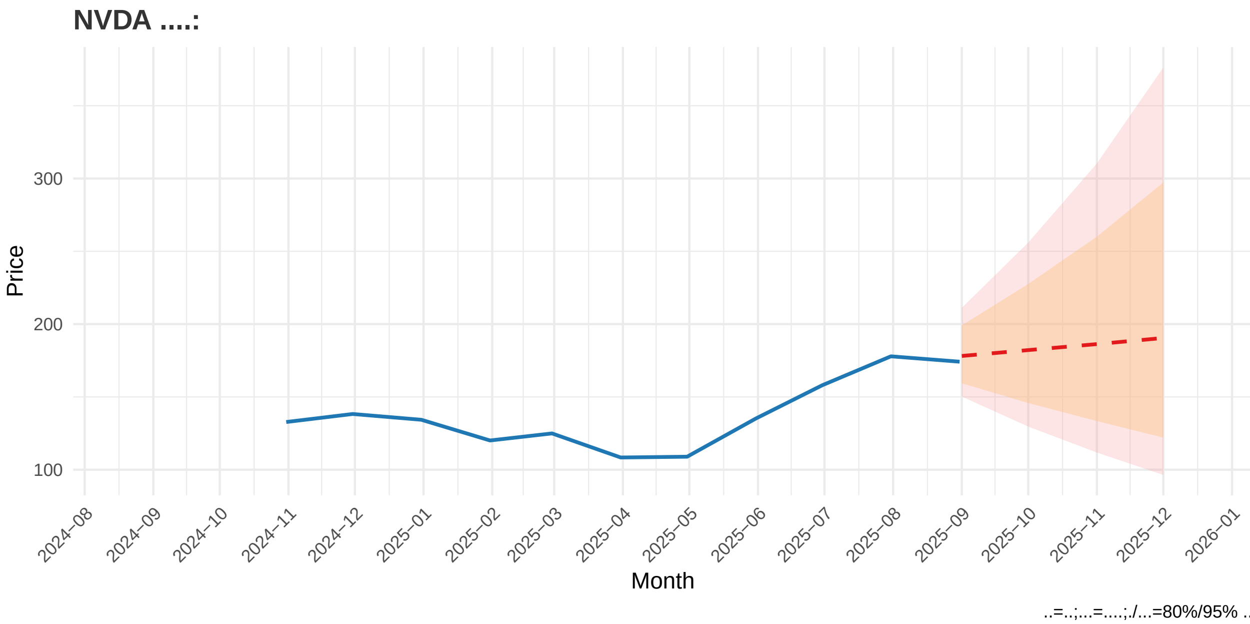
```
library(lubridate)
```

```
Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```
hist_df <- data %>%
  mutate(MonthEnd = as.Date(MonthEnd)) %>%
  filter(MonthEnd >= as.Date("2024-08-01"), MonthEnd <= as.Date("2025-08-31")) %>%
  select(MonthEnd, NVDA)

ret_mean <- as.numeric(fc$mean)
ret_lo80 <- as.numeric(fc$lower[, "80%"])
ret_hi80 <- as.numeric(fc$upper[, "80%"])
ret_lo95 <- as.numeric(fc$lower[, "95%"])
ret_hi95 <- as.numeric(fc$upper[, "95%"])

#
last_price <- as.numeric(tail(data$NVDA, 1))

px_mean <- last_price * cumprod(exp(ret_mean))
px_lo80 <- last_price * cumprod(exp(ret_lo80))
px_hi80 <- last_price * cumprod(exp(ret_hi80))
px_lo95 <- last_price * cumprod(exp(ret_lo95))
px_hi95 <- last_price * cumprod(exp(ret_hi95))

#          2025/08 →
start_month <- floor_date(max(as.Date(data$MonthEnd)), "month") + months(1)
future_months <- seq.Date(from = start_month, by = "month", length.out = length(ret_mean

fc_df <- data.frame(
  MonthEnd = future_months,
  px_mean  = px_mean,
  px_lo80  = px_lo80,
  px_hi80  = px_hi80,
  px_lo95  = px_lo95,
```

```r
  px_hi95   = px_hi95
)

# ==== 2)     +    +    ====
g <- ggplot() +
  #
  geom_line(data = hist_df, aes(MonthEnd, NVDA),
            linewidth = 1, color = "#1f78b4") +
  # 95%
  geom_ribbon(
    data = fc_df,
    aes(x = MonthEnd, ymin = px_lo95, ymax = px_hi95),
    fill = "#fb9a99", alpha = 0.25
  ) +
  # 80%
  geom_ribbon(
    data = fc_df,
    aes(x = MonthEnd, ymin = px_lo80, ymax = px_hi80),
    fill = "#fdbf6f", alpha = 0.35
  ) +
  #
  geom_line(
    data = fc_df,
    aes(MonthEnd, px_mean),
    linetype = "dashed", linewidth = 1, color = "#e31a1c"
  ) +
  scale_x_date(
    date_labels = "%Y-%m",
    breaks = "1 month",
    limits = c(as.Date("2024-08-01"), as.Date("2025-12-31")),
    expand = expansion(mult = c(0.01, 0.02))
  ) +
  scale_y_continuous(labels = label_number(big.mark = ",")) +
  labs(
    title = "NVDA    ",
    x = "Month", y = "Price",
    caption = "         /   80%/95%   "
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold", color = "#333333"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```

```r
print(g)
```



7 Conclision

由於資料只有近一年的，導致ARIMAX的參數估計不是很穩定，因此分析結果只能當參考之用，

上面提到的ARIMA(0,0,0)為簡單回歸，預測輝達股價會因為SOXX的股價影響， β 為 1.025308 (正相關) 可以得知當SOXX股票漲，輝達就會漲。