

Prediction

Youlan Shen

2023-04-29

Task 4: Prediction Using Estimated Model Parameters

```
library(tidyverse)
library(dplyr)
```

load data

```
load("hurricane.RData")
data <- hurricane %>% ungroup()
load("test.RData")
B <- test$B
gamma <- test$gamma
```

```
# get XYD
Y <- as.matrix(data$Wind.kt)
D <- data %>% mutate(intercept = 1) %>% dplyr::select(intercept, Wind_prev,
                                                    Lat_change, Long_change,
                                                    Wind_change) %>%
  as.matrix()
X <- data %>%
  mutate(index = 1:nrow(data)) %>%
  group_by(index) %>%
  dplyr::select(Month, Season, Nature) %>%
  mutate(n = 1) %>%
  pivot_wider(names_from = Month, values_from = n, values_fill = 0) %>%
  mutate(n = 1) %>%
  pivot_wider(names_from = Nature, values_from = n, values_fill = 0) %>%
  ungroup() %>%
  dplyr::select(April, May, June, July, August, September, October, November,
                December, Season, ET,NR,SS,TS) %>%
  as.matrix()
```

```
## Adding missing grouping variables: 'index'
```

Get Estimated Model Parameters Using MCMC Train

```

# get parameter estimates function
get_estimate <- function(null_matrix, iter, MCMCchain){
  for (i in iter) {
    null_matrix <- null_matrix + MCMCchain[[i]]
  }
  estimate <- null_matrix / length(iter)
  estimate
}

# set iter, we only use the last 3000 of MCMC to get posterior mean
iter <- 3001:6000
# set Null matrix
beta_estimate <- matrix(data = 0, nrow = 5, ncol = 699)
gamma_estimate <- matrix(data = 0, nrow = 14, ncol = 1)
beta_estimate <- get_estimate(beta_estimate, iter, B)
gamma_estimate <- get_estimate(gamma_estimate, iter, gamma)

```

Get Prediction

```

X_with_id <- cbind(as.numeric(factor(data$ID)), X)
#number of hurricanes
H <- max(X_with_id[,1])
ith_hurricane_idx <- 1:H |>
  map(\(i) which(X_with_id[,1] == i))
# calculate Y estimates
N <- length(Y)
mu_H <- rep(NA, N)
for (i in 1:H) {
  curr_hurricane_idx <- ith_hurricane_idx[[i]]
  mu_i <- D[curr_hurricane_idx,,drop = FALSE] %*%
    beta_estimate[,i,drop = FALSE]
  mu_H[curr_hurricane_idx] <- mu_i
}
mu_H <- as.matrix(mu_H)
Wind_pred <- mu_H + X %*% gamma_estimate

```

Analysis on Prediction

```

hurricane["Wind_pred"] <- Wind_pred
hurricane <- hurricane %>%
  dplyr::select(ID, Season, Month, Nature, Date, Wind.kt, Wind_pred)
hurricane %>%
  mutate(Pred_E = Wind.kt - Wind_pred,
         Pred_E_Sq = Pred_E * Pred_E) %>%
  group_by(ID) %>%
  summarise(nobs = n(),
            Pred_SE = sqrt(sum(Pred_E_Sq))) %>%
  arrange(Pred_SE) %>%
  top_n(10) %>%
  knitr::kable(

```

```

caption = "Prediction Standard Error of Wind Speed in Ascending Order",
col.names = c("Hurricane ID", "Number of Observation", "Prediction SE"),
digits = 3
)

```

```
## Selecting by Pred_SE
```

Table 1: Prediction Standard Error of Wind Speed in Ascending Order

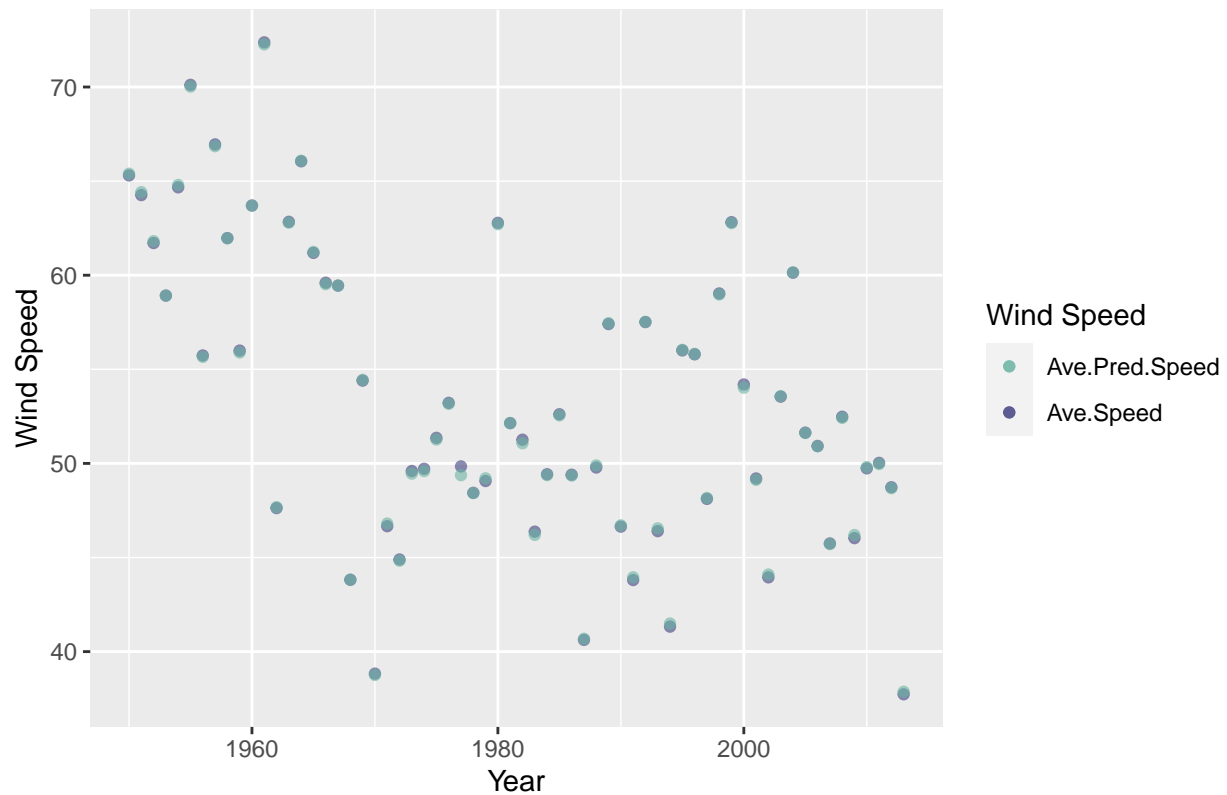
Hurricane ID	Number of Observation	Prediction SE
ETHEL.1960	12	68.844
AUDREY.1957	15	68.851
BEULAH.1967	68	70.419
IRIS.2001	20	71.388
IVAN.2004	92	72.783
JANET.1955	33	73.912
INEZ.1966	79	74.201
FELIX.2007	26	75.803
DENNIS.2005	57	79.094
EMILY.2005	43	80.648

```

hurricane %>%
  group_by(Season) %>%
  summarise(Ave.Speed = mean(Wind.kt),
            Ave.Pred.Speed = mean(Wind_pred)) %>%
  ggplot(aes(x = Season)) +
  geom_point(aes(y = Ave.Speed, color = "Ave.Speed"), alpha = 0.6) +
  geom_point(aes(y = Ave.Pred.Speed, color = "Ave.Pred.Speed"), alpha = 0.6) +
  scale_color_manual(name = "Wind Speed",
                    values = c("Ave.Speed" = "#404080",
                              "Ave.Pred.Speed" = "#69b3a2")) +
  labs(title = "Figure: Actual Wind Speed V.S. Prediction Wind Speed By Year",
       x = "Year",
       y = "Wind Speed")

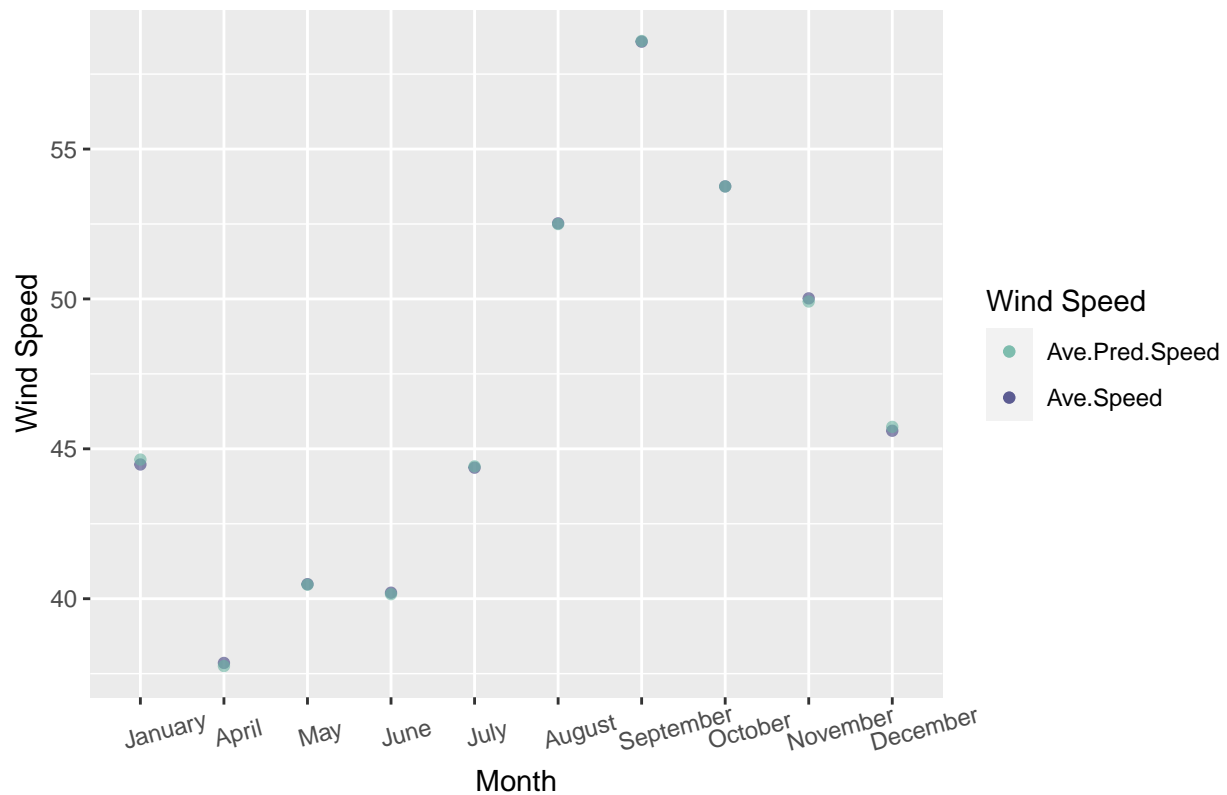
```

Figure: Actual Wind Speed V.S. Prediction Wind Speed By Year



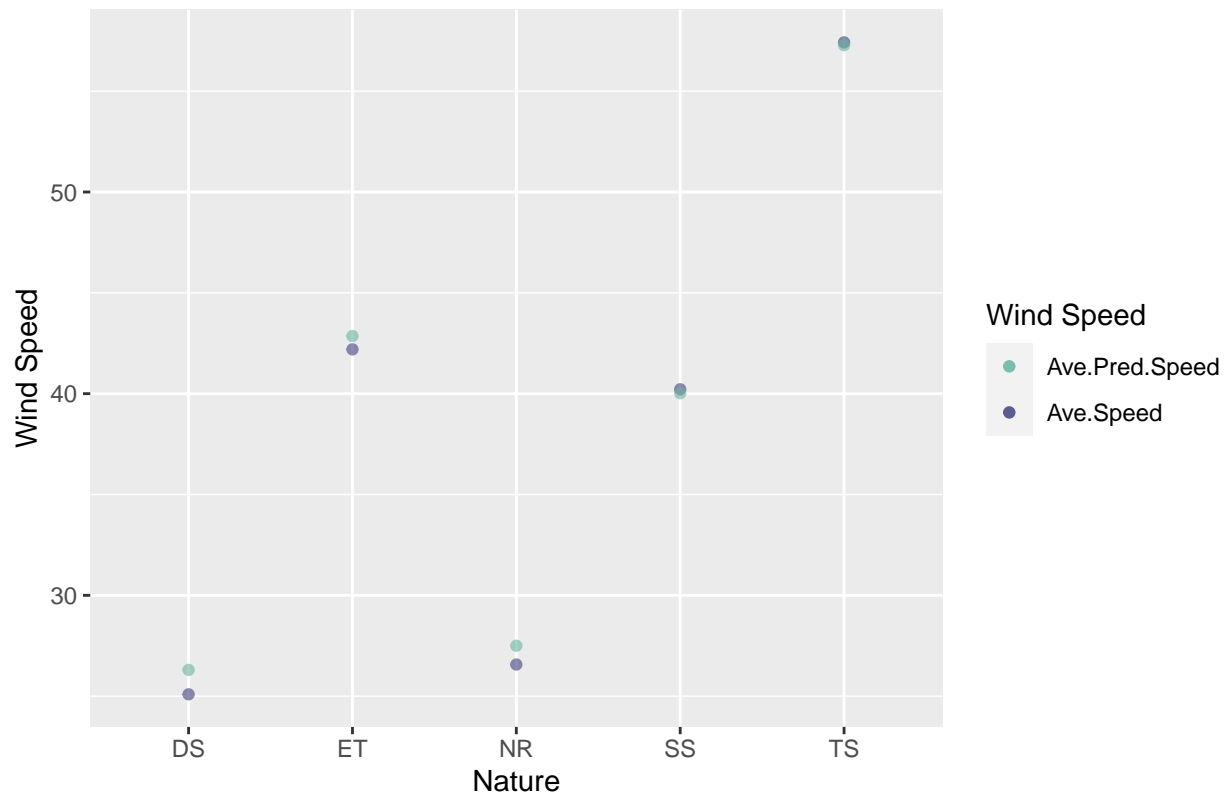
```
hurricane %>%
  group_by(Month) %>%
  summarise(Ave.Speed = mean(Wind.kt),
            Ave.Pred.Speed = mean(Wind_pred)) %>%
  ggplot(aes(x = Month)) +
  geom_point(aes(y = Ave.Speed, color = "Ave.Speed"), alpha = 0.6) +
  geom_point(aes(y = Ave.Pred.Speed, color = "Ave.Pred.Speed"), alpha = 0.6) +
  scale_color_manual(name = "Wind Speed",
                    values = c("Ave.Speed" = "#404080",
                              "Ave.Pred.Speed" = "#69b3a2")) +
  labs(title = "Figure: Actual Wind Speed V.S. Prediction Wind Speed By Month",
       x = "Month",
       y = "Wind Speed") +
  theme(axis.text.x = element_text(angle = 15, vjust = 0.5, hjust = 0.2))
```

Figure: Actual Wind Speed V.S. Prediction Wind Speed By Month



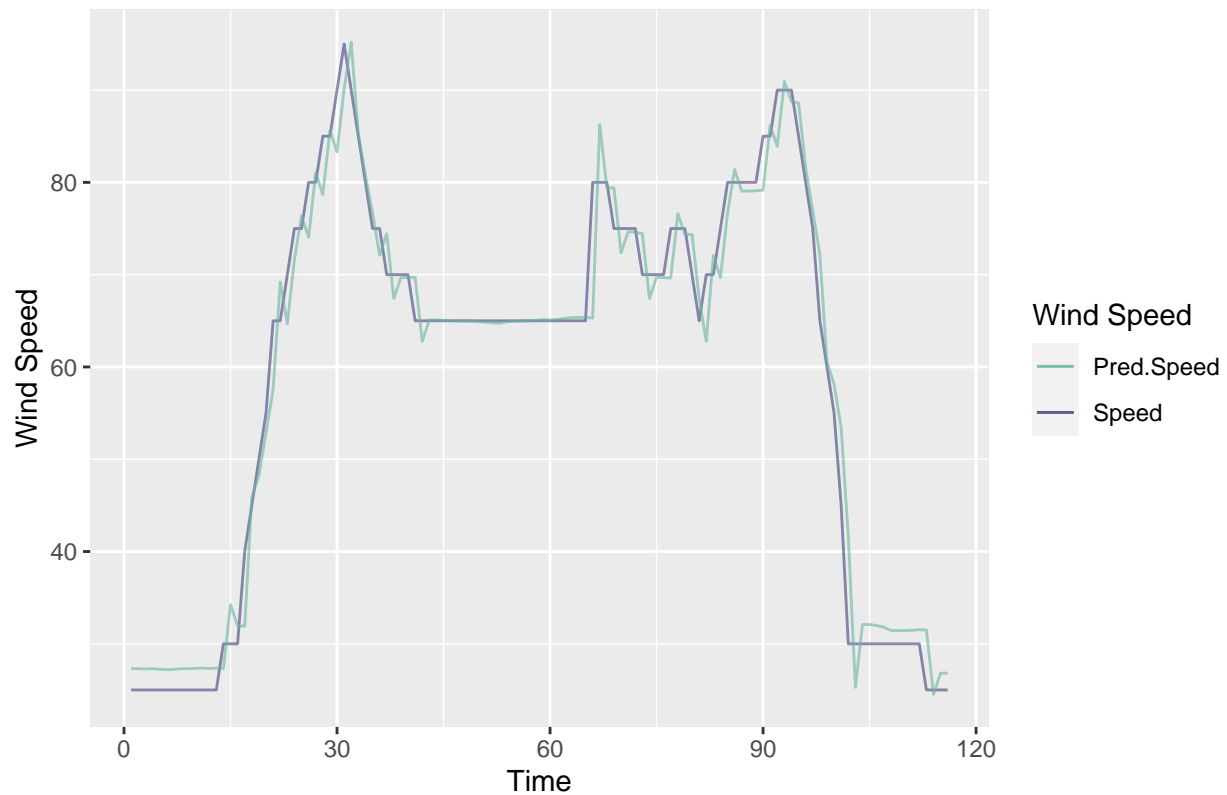
```
hurricane %>%
  group_by(Nature) %>%
  summarise(Ave.Speed = mean(Wind.kt),
            Ave.Pred.Speed = mean(Wind_pred)) %>%
  ggplot(aes(x = Nature)) +
  geom_point(aes(y = Ave.Speed, color = "Ave.Speed"), alpha = 0.6) +
  geom_point(aes(y = Ave.Pred.Speed, color = "Ave.Pred.Speed"), alpha = 0.6) +
  scale_color_manual(name = "Wind Speed",
                    values = c("Ave.Speed" = "#404080",
                              "Ave.Pred.Speed" = "#69b3a2")) +
  labs(title = "Figure: Actual Wind Speed V.S. Prediction Wind Speed By Nature",
       x = "Nature",
       y = "Wind Speed")
```

Figure: Actual Wind Speed V.S. Prediction Wind Speed By Nature



```
hurricane %>%
  filter(ID == "GINGER.1971") %>%
  mutate(Obs = 1:116) %>%
  ggplot(aes(x = Obs)) +
    geom_line(aes(y = Wind.kt, color = "Speed"), alpha = 0.6) +
    geom_line(aes(y = Wind_pred, color = "Pred.Speed"), alpha = 0.6) +
    scale_color_manual(name = "Wind Speed",
                       values = c("Speed" = "#404080",
                                   "Pred.Speed" = "#69b3a2")) +
    labs(title = "Figure: Actual Wind Speed V.S. Prediction Wind Speed For GINGER.1971",
         x = "Time",
         y = "Wind Speed")
```

Figure: Actual Wind Speed V.S. Prediction Wind Speed For GINGER.1971



```
hurricane %>%
  mutate(Pred_E = Wind.kt - Wind_pred,
         Pred_E_Sq = Pred_E * Pred_E) %>%
  group_by(ID) %>%
  summarise(n = n(),
            Pred_SE = sqrt(sum(Pred_E_Sq))) %>%
  mutate(nobs = ifelse(n > 50, "More Observation", "Fewer Observation")) %>%
  group_by(nobs) %>%
  ggplot(aes(x = nobs, y = Pred_SE)) +
  geom_boxplot(fill = "#69b3a2", alpha = 0.6) +
  labs(title = "Figure: Prediction Standard Error of Wind Speed by Number of Observation",
       x = "Observation",
       y = "Prediction Standard Error")
```

Figure: Prediction Standard Error of Wind Speed by Number of Observation

