Simple Linear Regresson of Building Dynamics

Edward J. Xu (edxu96@outlook.com)

2019-09-15

```
library(tidyverse)
library(broom)
library(magrittr)
library(forecast)
library(ggfortify)
```

Read the data about heat load in Soenderborg. The model for house 3 is to be modeled, so only the heat load from house 3 is selected. Also, filter the data in winter and store them in `ti_winter`.

```
ti <-
  read_csv("~/GitHub/tidynamics/data/soenderborg_2day.csv") %>%
  mutate(
    "is_win" = as.POSIXlt("2010-11-01") <= as.POSIXlt(.$t) &
      as.POSIXlt(.$t) < as.POSIXlt("2011-02-01")
  ) %>%
  select(t, P3, Te, G, Ws, is_win) %>%
  drop_na()
```

1, Simple Linear Regression

```
li_mod <- list(
  "entire" = lm(P3 ~ Te, ti),
  "winter" = lm(P3 ~ Te, filter(ti, is_win))
)
```

```
li_mod %>%
  map_df(tidy, .id = "period")
#> # A tibble: 4 x 6
#>   period term          estimate std.error statistic  p.value
#>   <chr>  <chr>            <dbl>     <dbl>     <dbl>    <dbl>
#> 1 entire (Intercept)     3625.      53.6      67.6 2.72e-109
#> 2 entire Te              -210.       4.82     -43.6 9.36e- 84
#> 3 winter (Intercept)     3751.      60.7      61.8 2.15e- 44
#> 4 winter Te              -217.      13.5      -16.1 4.31e- 20
```

```
li_mod %>%
  map_df(glance, .id = "period")
#> # A tibble: 2 x 12
```

```
#>   period r.squared adj.r.squared sigma statistic  p.value   df logLik
#>   <chr>      <dbl>        <dbl> <dbl>     <dbl>    <dbl> <int>  <dbl>
#> 1 entire     0.931        0.931  453.     1903. 9.36e-84     2 -1076.
#> 2 winter     0.855        0.852  411.      260. 4.31e-20     2  -341.
#> # … with 4 more variables: AIC <dbl>, BIC <dbl>, deviance <dbl>,
#> #   df.residual <int>
```

The P-values of coefficients for intercept and `Te` are too small, so neither of them is a qualified model.

1.1, Residual Analysis

```
autoplot(li_mod$entire, data = ti, colour = "is_win")
```
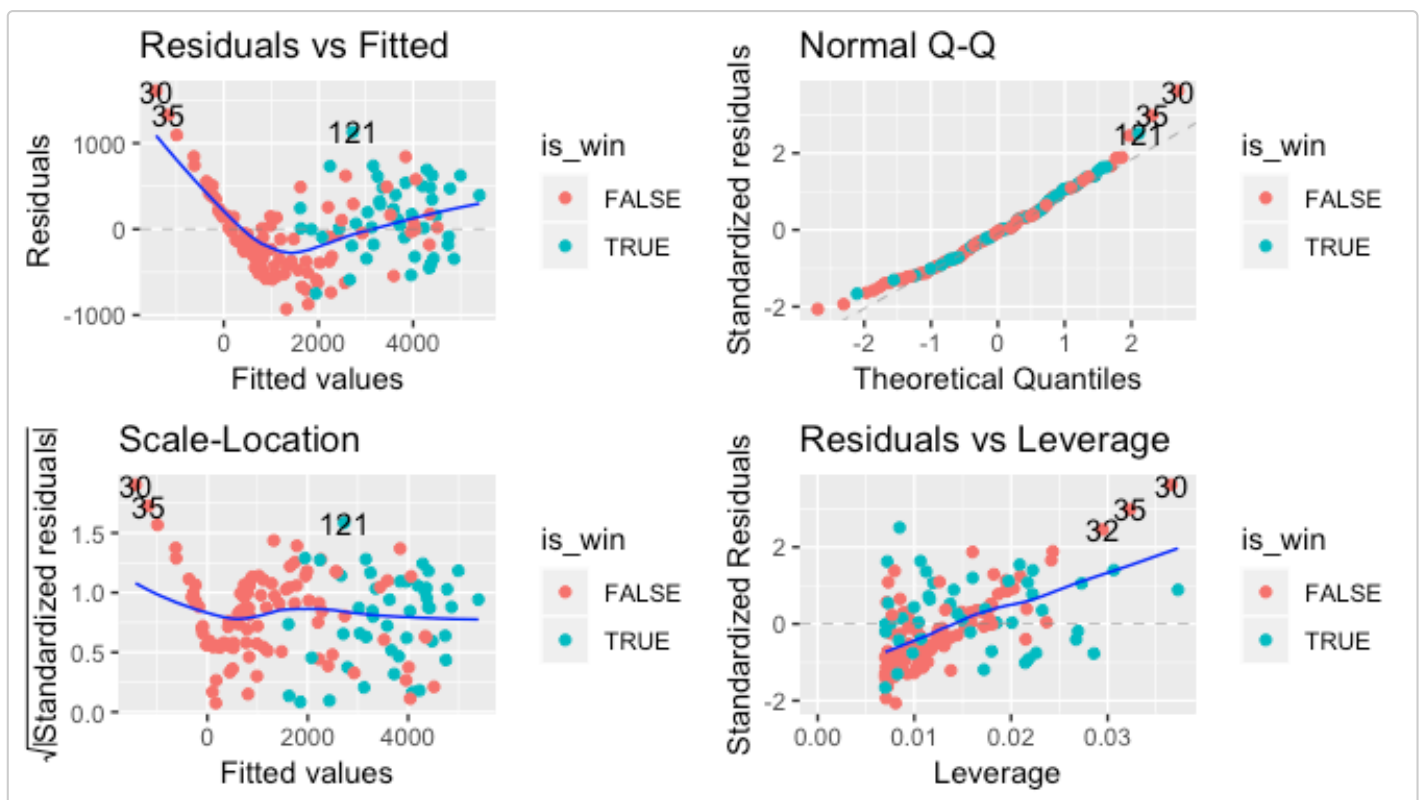

Linear Reg of Heating Load (y) and Ext Temp (x) in Entire Period

According to the two figures in left panel, the variability of the residuals suggests that the variance of heat load is related to its mean, which violates the costant variance assumption. There are many large residuals when the heat load is smaller than 0 during spring, summer and fall. Furthermore, the lower right panel shows there are some observations with large leverages affecting the fitted model.

```
autoplot(li_mod$winter)
```
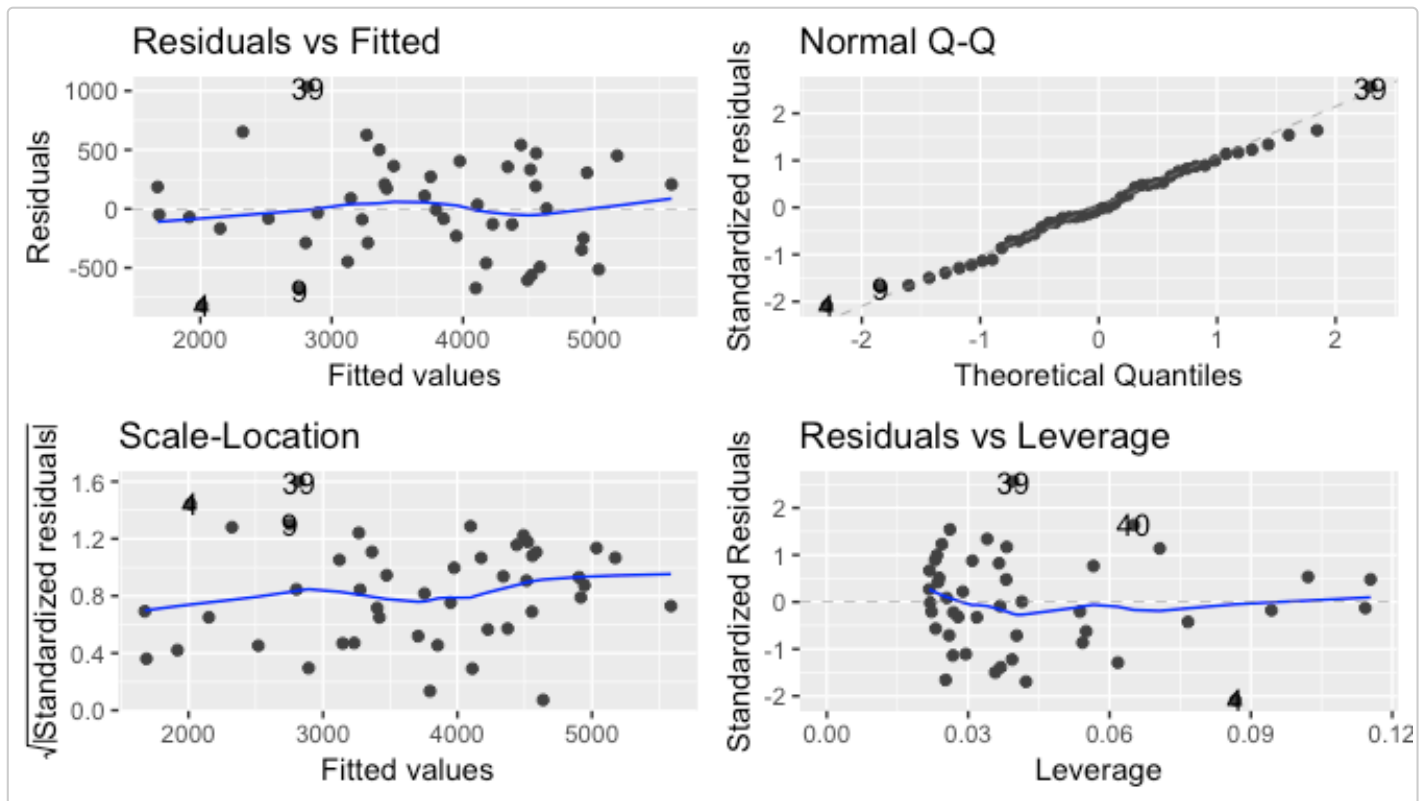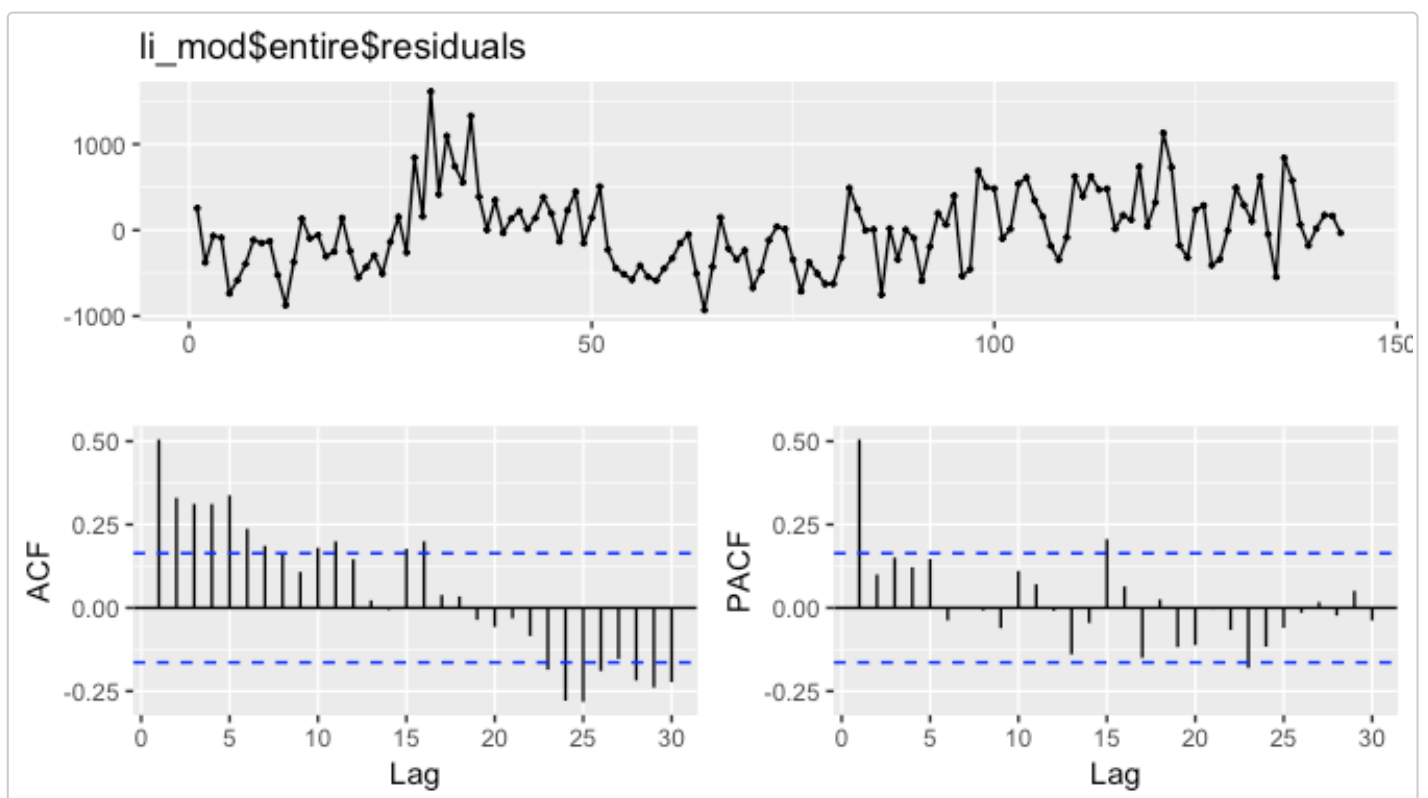
Linear Reg of Heat Load (y) and Ext Temp (x) in Winter

However, the two figures in left panel shows that the model for heat load in winter period fits quite well. The qqplot behaves much better as well. There is no observations with large leverages and large residuals.
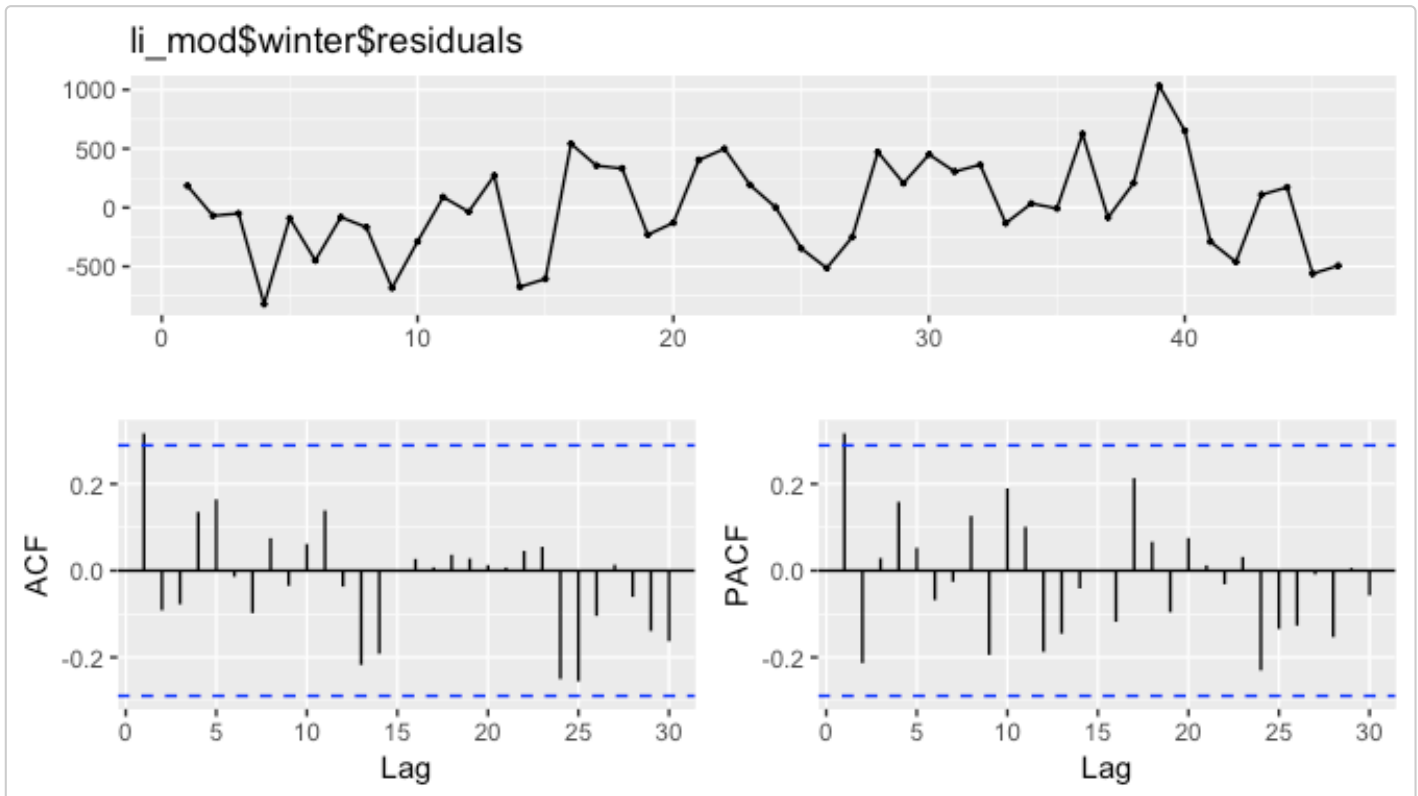
1.2, Validation of i.i.d. Assumption

```
forecast::ggtsdisplay(li_mod$entire$residuals, lag.max = 30)
```

ACF of Residuals from Linear Reg Model `lm_winter_3`

```
forecast::ggtsdisplay(li_mod$winter$residuals, lag.max = 30)
```


ACF of Residuals from Linear Reg Model `lm_3`

The residuals from the model fitted with data from entire period show correlation in both ACF and PACF, while those in winter period show insignificant correlation.