STA 642 - project

Jae Hyun Lee, jl914

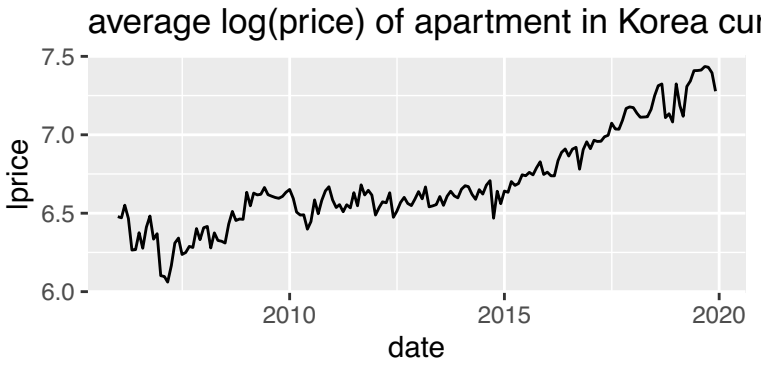
23 March, 2020

1. The topic and outline

Topic

I choose forecasting the market price of apartment in Korea as a topic of project. For a long time, real estates have been most popular investment item in Korea. Moreover, as concentration of city in Korea becomes severe, the market price of apartment gets even higher. These facts motivate me to develop model which can technically and precisely forecast future price of apartments. We can find increasing trend of average price of apartment.

```
ggplot() +
  geom_line(data = m_price, mapping = aes(x = date, y = lprice)) +
  labs(title = "average log(price) of apartment in Korea currency")
```



Data

The Ministry of Land, Infrastructure, and Transport in Korea provide actual transaction of apartment data in "http://rtdown.molit.go.kr/". We can use this site and data freely. This link provides transaction of apartment from 2006 to the most recent time point. I downloaded transaction data of "Gangnam" district which is considered as the most expensive area from 2006 to 2019. This data includes columns as follow: `Official address`, `District`, `District in large unit`, `District in small unit`, `Name of apartment`, `Size`, `Year and Month`, `Day`, `Price`, `Floor`, `Time apartment built`, and `Address in Road name`.

Current Goal of analysis

My current goal is developing Multivariate dynamic linear model which focus on Dynamic regression. I want to develop model whose linear slope evolve as time pass so that we can reflect time changing factors.

2. Ideas from course materials and anticipating model

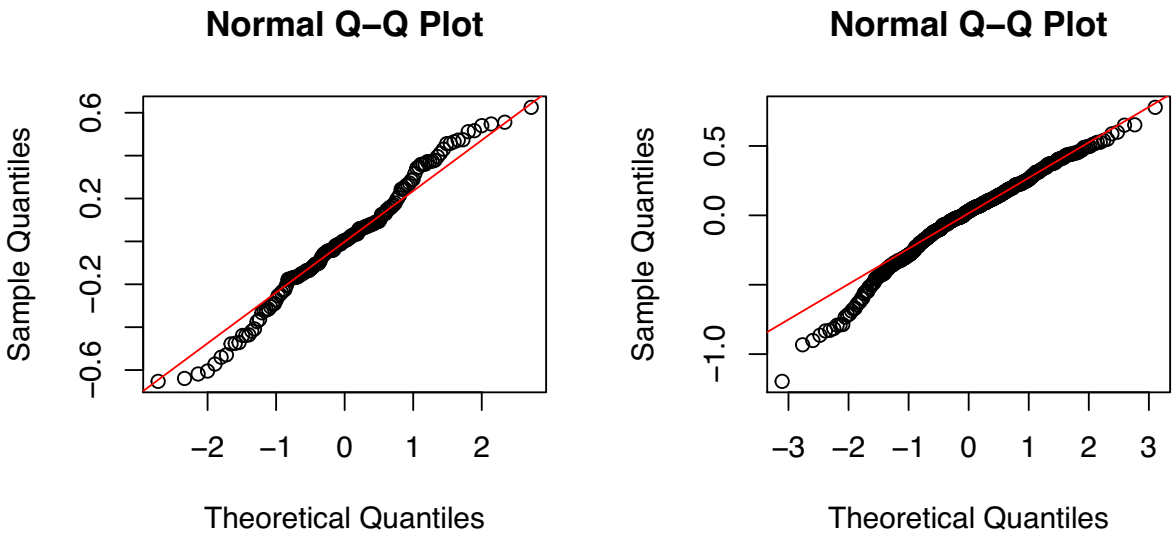
As stated above, I will bring Dynamic regression model in chapter 3 and in P&W. In this project, I want to use dynamic linear model as follow:

$$y_t = F_t\beta_t + v_t \quad v_t \sim N(0, \nu_t I_n)$$
$$\beta_t = G_t\beta_{t-1} + \epsilon_t \quad \epsilon_t \sim N(0, W_t)$$

where target variable y_t is log-transformed price vector, F_t is $\begin{pmatrix} 1 & X_{1,t} & X_{2,t} & X_{3,t} \end{pmatrix}$ which is predictor matrix of time t and β_t = coefficient of intercept and predictors so that I can use time varying coefficients to predict appropriate price for apartment.

For the target variable, I decided to log transform original target variable to satisfy normality assumption of model. As a result, we can see that normality assumptions for each time point is not severely violated.

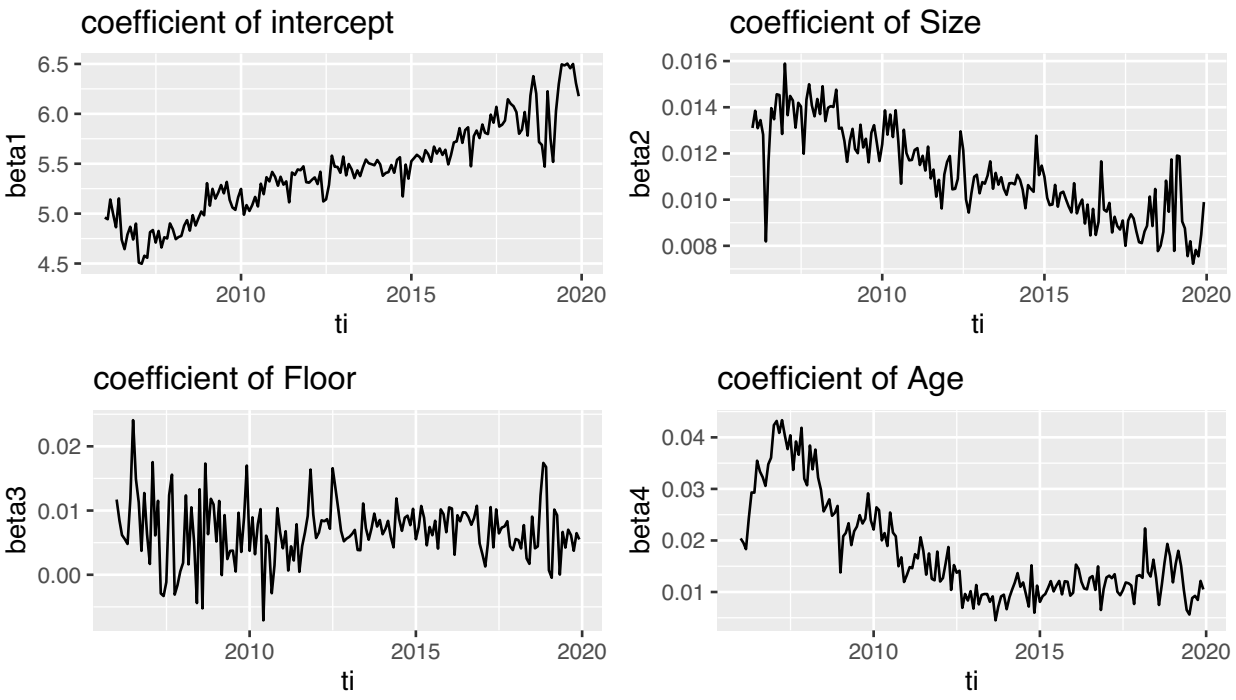
```
par(mfrow = c(1,2))
qqnorm(resid[[sam[1]]])
  qqline(resid[[sam[1]]],col = "red")
qqnorm(resid[[sam[2]]])
  qqline(resid[[sam[2]]],col = "red")
```



Evolution matrix G_t is not determined yet, but first attempt would be AR(1) model. It will be extended to AR(p) or TAVR(p). Since linear regression assumes that errors are indepedent, I will apply that assumption in my model too. In evolving and updating process, I will use result from course materials and P&W.

I have confirmed that my first inspiration time varying slope is reasonable hen we investigate βs of each time point. At these plots, we can find that we should fit time varying coefficients rather than constants.

```
grid.arrange(grobs = plot_list,nrow = 2)
```



we can also find that estimated variance of y_t becomes very unstable, thus I plan to use volatile model and large discount factor for $V(v_t) = \nu_t$ to reflect this phenomenon.

```
plot_sigma
```