Homework 3 of *Essentials of Regression Analysis Using R*

Keio University

Due: Friday January 19th 2024 in class

We keep the same dataset *SRE.csv* as in the previous take-home exam. We recall that the six variables are:

- *Price* corresponds to the price of the transaction (in US dollars).
- *Priceop* corresponds to the price (in US dollars) of the best opposite ask/bid.
- *Trade* is the trade indicator, i.e. equal to 1 if the transaction is buyer initiated or equal to -1 if the transaction is seller initiated.
- *Time* corresponds to the transaction time (in second).
- *Volume* corresponds to the volume (in number of shares) of the transaction.
- *Depth* corresponds to the depth (in number of shares) in the limit order book.

For any variable $V \in \{\text{Price}, \text{Priceop}, \text{Trade}, \text{Time}, \text{Volume}, \text{Depth}\}$, we have $n = 15127$ (intraday) observations denoted V_i.

We define the (positive) variable $\text{Spread}_i := \mid \text{Price}_i - \text{Priceop}_i \mid$ and the signed spread as $\text{SSpread}_i := \text{Trade}_i \times \text{Spread}_i$. We also define $\text{SVolume}_i := \text{Trade}_i \times \text{Volume}_i$ and $\text{SDepth}_i := \text{Trade}_i \times \text{Depth}_i$.

For any variable $V = (V_1, \cdots, V_n)$, we define the lagged differences as $\Delta V := (V_2 - V_1, \cdots, V_n - V_{n-1})$. For example, if $V = (1, 2, 4)$, then $\Delta V = (2 - 1, 4 - 2)$, i.e.

$\Delta V = (1, 2)$. Bear in mind that the size of V is n and ΔV is $n-1$. Finally, we define $\text{S}\Delta\text{Time}_i := \text{Trade}_{i+1} \times \Delta\text{Time}_i$ for $i = 1, \cdots, n-1$. We aim to fit the following linear model:

$$\begin{aligned}\Delta\text{Price} &= \theta_0 + \theta_1 \Delta\text{Trade} + \theta_2 \Delta\text{SSpread} + \theta_3 \Delta(S\Delta\text{Time}) + \theta_4 \Delta\text{SVolume} \\ &\quad + \theta_5 \Delta\text{SDepth} \end{aligned} \tag{1}$$

1 Pre-process of the data

1. Do you think there is any adequate transformation of the Price variable you could make (such as the square of Price variable). If so, please implement such transformation. If not, do not change anything in the variable Price.

2 Analysis

1. As for Homework 2, we also consider the linear model incorporating *Time* as a sixth regressor:

$$\begin{aligned}\Delta\text{Price} &= \theta_0 + \theta_1 \Delta\text{Trade} + \theta_2 \Delta\text{SSpread} + \theta_3 \Delta(S\Delta\text{Time}) + \theta_4 \Delta\text{SVolume} \\ &\quad + \theta_5 \Delta\text{SDepth} + \theta_6 \text{Time}. \end{aligned} \tag{2}$$

   Implement an hypothesis test to compare models and see whether the fit of (2) significantly improves the fit of (1). Interpret the results of the test statistics.

2. Fit two new models:

$$\begin{aligned}\Delta\text{Price} &= \theta_0 + \theta_1 \Delta\text{Trade} + \theta_2 \Delta\text{SSpread} + \theta_3 \Delta(S\Delta\text{Time}) + \theta_4 \Delta\text{SVolume} \\ &\quad + \theta_5 \Delta\text{SDepth} + \theta_6 (\Delta\text{SSpread})^2 \end{aligned} \tag{3}$$

   and

$$\begin{aligned}\Delta\text{Price} &= \theta_0 + \theta_1 \Delta\text{Trade} + \theta_2 \Delta\text{SSpread} + \theta_3 \Delta(S\Delta\text{Time}) + \theta_4 \Delta\text{SVolume} \\ &\quad + \theta_5 \Delta\text{SDepth} + \theta_6 (\Delta\text{SSpread})^2 + \theta_7 (\Delta\text{SSpread})^3. \end{aligned} \tag{4}$$

   Provide the tables, give an interpretation and implement an hypothesis test to compare models to see whether both models improve the fit of (1) or not.

3. Make predictions based on model (1) at the point Intercept = 0, ΔTrade = 2, ΔSSpread = 0, $\Delta(S\Delta\text{Time}) = 0.1$, ΔSVolume = 100, ΔSDepth = 100, and provide two different confidence intervals (the confidence interval for "prediction of a future value" and "prediction of the mean response"). Interpret the results.

4. Implement a test of constant variance of the residuals in (1). Illustrate with a plot and interpret the results. Discuss the implications of the results regarding inference.

5. Same question with possible correlation between the residuals.

6. What do you think about the normal assumption of the errors? No formal test is required in this part. Discuss the implications of the results regarding inference.