Homework #4 - Glosten-Harris model

FE-570

April 2, 2022

Glosten-Harris model. The Glosten-Harris model is an improved version of the Roll model which takes into account price impact of trades.

The trade price p_t is given by the sum of the *efficient price* m_t, the trading cost cd_t with d_t the trade sign (± 1 for buy/sell), and a price impact term λx_t

$$p_t = m_t + cd_t + \lambda x_t\,. \tag{1}$$

x_t are *signed* trade sizes: positive for buy, negative for sell.

The efficient price m_t evolves according to the process

$$m_t = m_{t-1} + \lambda x_{t-1} + u_t \tag{2}$$

Here u_t are iid random variables with mean zero and variance σ_u^2, just like in the Roll model. The new term λx_{t-1} is the price impact contribution due to the previous trade.

Calibrate the λ, c parameters of this model on the JPM tick level dataset in *taqdata_JPM_20210113_ESTMktHrs.RData*.

For this analysis it is important to use only trades executed on one exchange, as different exchanges have different price impact. For this problem use only the trades executed on ADF. Also exclude from the analysis large trades at the start and end of the trading session. For example keep only trades with time stamps between 10:00am - 12:00pm.

Use *getTradeDirection* to estimate the trade signs d_t.

Hints

First, we need the signed trade sizes $x_t = d_t|x_t|$, where d_t are the trade indicators and $|x_t|$ are the absolute values of the trade size which are available in the TAQ data as `tqdata$SIZE`. We get the trade indicators using the `getTradeDirection(tqdata)` which is the implementation of the Lee-Ready mid-point criterion.

The calibration proceeds in two steps.

Step 1. Determine λ. This is done using Equation (2) where we assume that the efficient price m_t is the same as the mid-price $\frac{1}{2}(a_t + b_t)$. The change of the efficient price from one trade to the next is

$$\Delta m_t = m_t - m_{t-1} = \lambda x_{t-1} + u_t \tag{3}$$

A linear regression of Δm_t with respect to $x(t-1)$ of the form $\Delta m_t = ax_{t-1} + b$ should give λ as the slope, and $b = \mathbb{E}[u_i]$ the average of the residuals.

Step 2. Determine c. This is done from Eq. (1) written as

$$p_t - m_t - \lambda x_t = cd_t \tag{4}$$

A linear fit of the left-hand side to the d_t should give c as the slope.