Math 584 (Math for Algo Trading), Spring 2022.

Homework 4

Due: Thu, Apr 7, 2022, NO LATER than 5pm.

1. In this exercise, you will estimate the coefficient of a linear price impact function and will deduce the risk-aversion of the (hypothetical) market maker from it.

   (a) (8 pts) Estimate the price impact coefficient λ, as discussed in Section 5 of Chapter 4, using the model:

$$\Delta S_t = \lambda \Delta \hat{V}_t + \hat{\sigma}\eta_t, \tag{1}$$

   where $\hat{V}$ is the **signed traded volume**, S is the midprice, η is a zero-mean unit-variance white noise, and $\Delta S_t = S_{t+\Delta t} - S_t$, $\Delta \hat{V}_t = \hat{V}_{t+\Delta t} - \hat{V}_t$. Use a reduced-frequency sample of $(\hat{V}, P)$, by choosing Δt that corresponds to one second (i.e., you consider the values of $(S, \hat{V})$ at the beginning of every second). Use the least-squares linear regression to estimate λ for each of the tickers MSFT and GOOG, using the data from Nov 3 through Nov 26, 2014, as a sample for each ticker.

   (b) (10 pts) Estimate the price impact coefficient λ via LOB. At each time t, consider the volumes $\{V_t^{a,i}, V_t^{b,i}\}_{i=1}^{10}$ on the ask and bid sides of the LOB, respectively, at the first 10 price levels $\{P_t^{a,i}, P_t^{b,i}\}_{i=1}^{10}$ (with $i = 1$ corresponding to the best price). Imagine a flat LOB with the same total volume (over the first 10 levels of LOs) on each side. The height of the latter (imaginary) LOB on the bid side is

$$H_t^b = \frac{1}{S_t - P_t^{b,10}} \sum_{i=1}^{10} V_t^{b,i},$$

   and its height on the ask side is

$$H_t^a = \frac{1}{P_t^{a,10} - S_t} \sum_{i=1}^{10} V_t^{a,i},$$

   where S_t is the midprice at time t. Notice that

$$\lambda_t = \frac{1}{2}\left(\frac{1}{H_t^b} + \frac{1}{H_t^a}\right)$$

   can be viewed as the (approximate) instantaneous price impact at time t.

   Estimate the average price impact λ as the sample mean of $\{\lambda_t\}$ for each of the tickers MSFT and GOOG, using the data from Nov 3 through Nov 26, 2014, as a sample for each ticker.

   Compare the resulting estimates of λ, for each ticker, to those obtained in part (a) and explain the difference. Which estimation method would you use when designing a trading strategy and why?

   (c) (8 pts) Recall that, in the utility indifference model (or in Grossman-Miller model with one market maker), the price impact coefficient is given by

$$\lambda = 2\gamma\sigma^2, \tag{2}$$

where γ measures the risk aversion of the market maker (the higher is γ, the more "afraid" the market maker is of the risk), and σ^2 is the variance of the "fundamental value" of the asset. The goal of parts (c) and (d) is to estimate γ.

While it is hard to measure the fundamental value of the asset, we can pretend that it is given by the midprice of the stock at the end of the next 5-minute time interval. Thus, to approximate σ^2, we denote by h the number of time stamps in a 5-minute time interval (for the sample data we have), and consider the associated 5-minute increments of the midprice, $S_{(i+1)h} - S_{ih}$. We approximate σ^2 by the sample variance of $\{S_{(i+1)h} - S_{ih}\}_i$.

Estimate σ^2 for each of the tickers MSFT and GOOG, using the data from Nov 3 through Nov 26, 2014, as a sample for each ticker. Compare the estimated values of σ^2 for each ticker. What does this comparison tell you about the behavior of the prices of these two stocks?

(d) (2 pts) Use the value of σ^2, estimated in part (c), and the two estimates of λ, obtained in parts (a) and (b), to produce four estimates of γ via (2). Comment on how these estimates compare to each other, across the two tickers and across the two estimation methods.

2. (30 pts) In this exercise, you implement and test the trading strategy described in Section 5.1 of Chapter 4, using the data for MSFT ticker on Nov 3, 2014, as your sample. The parameters' values are as follows (using the notation of Chapter 4):

$$q = 100, \quad \text{Trading time range: 10am–3:30pm},$$

$$\text{Length of estimation window: 1 minute},$$

$$\text{Length of prediction interval: 10 seconds},$$

$$tP^* = 0.00999, \quad lP^* = 0.00999, \quad p_1^* = 0.01, \quad p_2^* = 0.05.$$

(The reason for the above choice of tP^* and lP^*, which are not given by integer numbers of cents, is to protect the algorithm against the numerical (rounding) errors, which arise when the data is transferred from a MatLab format to Python.)

Compute the PnL of this strategy for every time stamp in the testing range 10:01am–3:30pm, so that, at any time t, you define the associated increment of the PnL process as the profit/loss from a roundtrip trade if this trade is terminated (i.e., the position is closed) at time t, and you set this increment to zero otherwise. Plot the resulting PnL process and compute the Sharpe ratio of its absolute returns. Annualize the Sharpe ratio by multiplying it by $\sqrt{250 \cdot n}$, where n is the length of the PnL process.