

TAQ Data workflow and EDA

HFT: Assignment 1

Julian Albert

University of Cape Town
Department of Statistics

August 15, 2019



Table of Contents

1 Introduction

2 Data

3 Exploratory Data Analysis

4 Order-Flow

5 Inter-Arrival Times

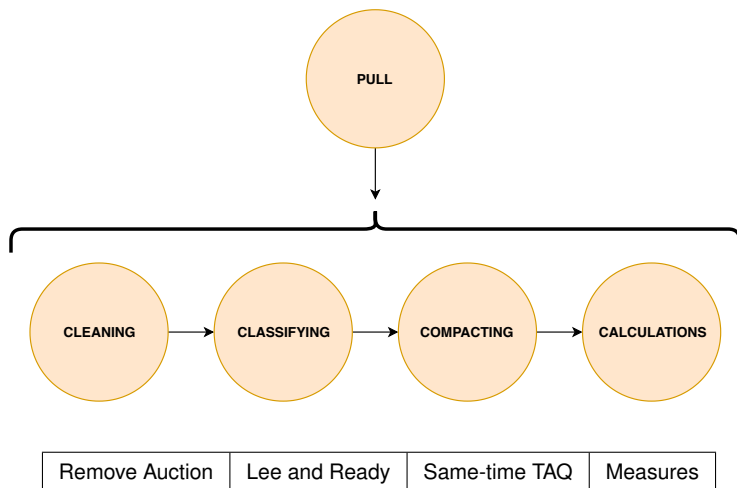
Introduction

Introduction

- Tasked with creating a trade and quote (TAQ) data workflow and performing some basic exploratory data analysis (EDA).
- TAQ data contains intraday trades and quotes providing insight into a securities supply and demand that drives price, an essential part of high-frequency data analysis.
- The TAQ data considers the 5 month period starting 2019-01-17 and ending 2019-06-19.
- We consider two securities, AGL: Anglo American PLC and NPN: Naspers Ltd.

Data

Data Processing



Exploratory Data Analysis

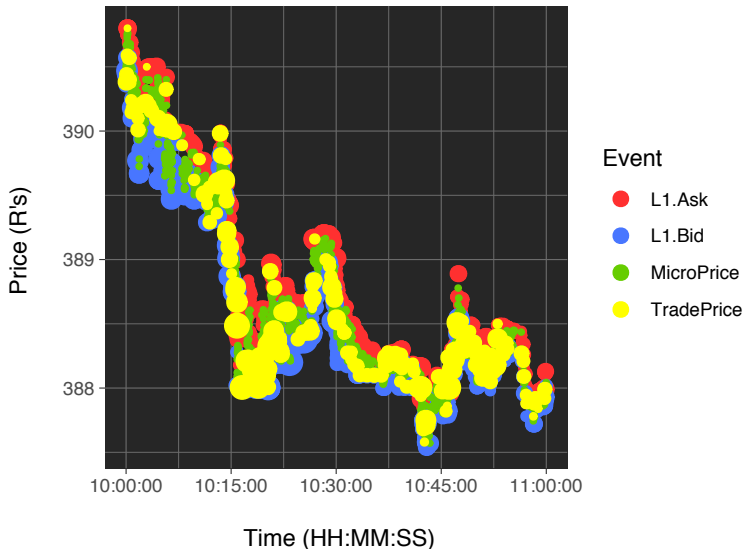
With the prepared dataset we can visualise an hour of trading on the two different stocks between 10h00 and 11h00, and 16h00 and 17h00. This visualisation is achieved by plotting

- the micro-price (green)
- the best-bid (Level 1) to buy price (blue)
- the best-ask (Level 1) on offer to sell price (red)
- the transaction price (yellow)

in calendar time with the order book event bubbles are weighted proportional to the volume on bid, on offer and transacted.

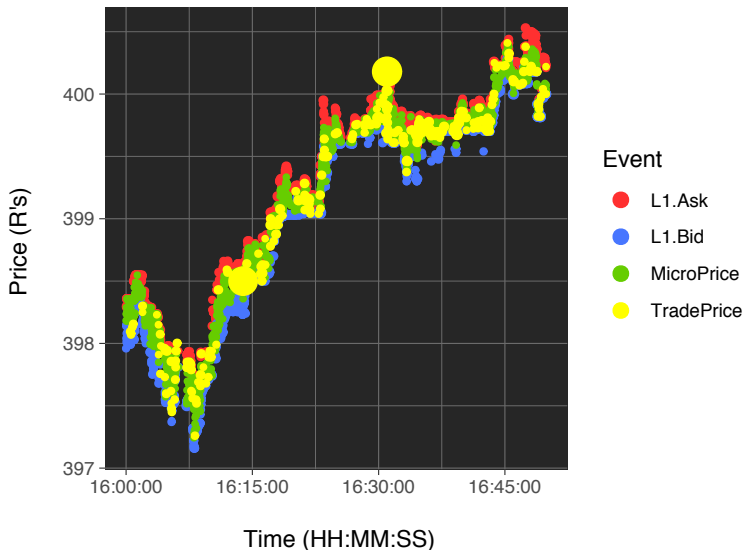
AGL: Anglo American PLC

Trading between 10h00 and 11h00 : 2019-04-17



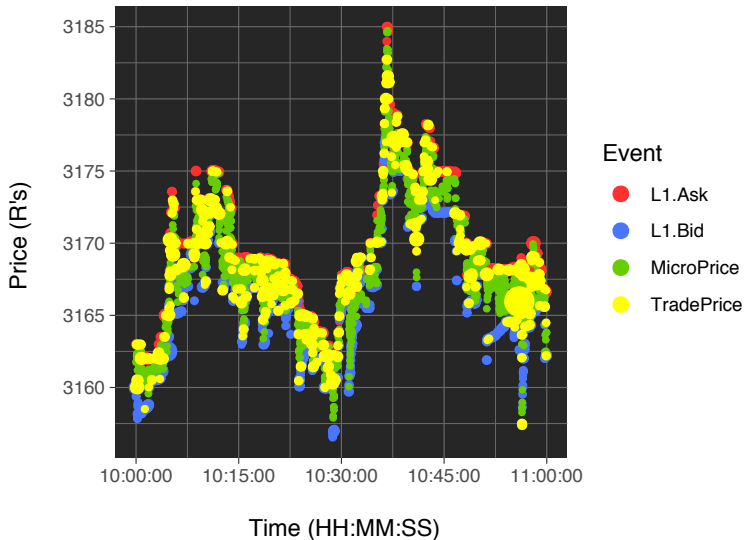
AGL: Anglo American PLC

Trading between 16h00 and 17h00 : 2019-04-23



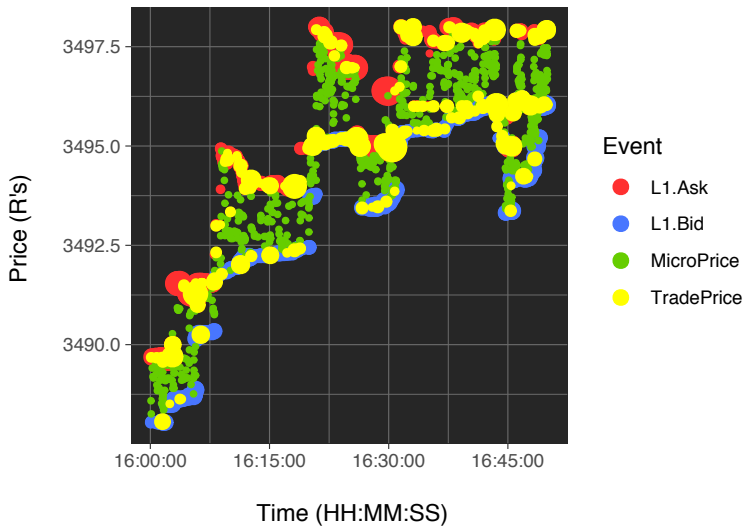
NPN: Naspers Ltd

Trading between 10h00 and 11h00 : 2019-03-04



NPN: Naspers Ltd

Trading between 16h00 and 17h00 : 2019-04-08



Order-Flow

Order flow can be defined as the the sequence/process which assumes value +1 for buyer initiated trades and -1 for seller initiated trades.

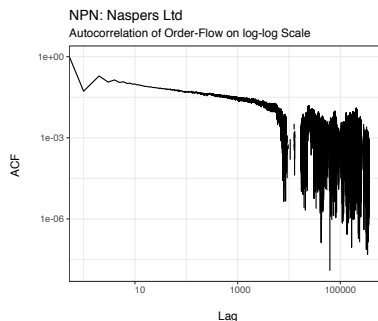
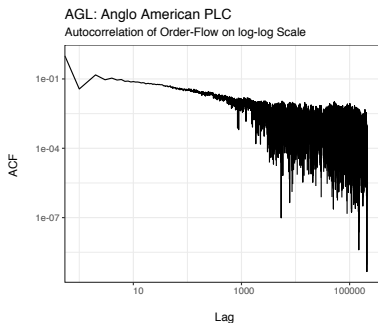


Figure: Auto-Correlation of Order Flow for Two Equities.

Inter-Arrival Times

Trade Frequency

We plot the histogram of inter-arrival times over the entire set of trading events for the 135 trading days. We can see that the amount of trading events are higher for Naspers Ltd compared to AGL PLC implying higher liquidity.

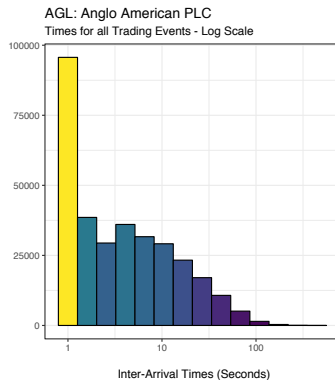
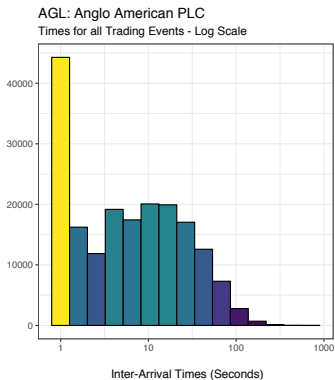


Figure: Frequency of Inter-Arrival Times on Log Scale.

Trade Frequency

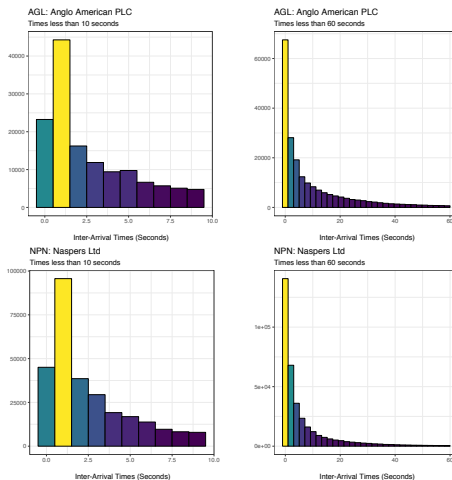


Figure: Frequency of Inter-Arrival Times less than 10 seconds and 1 minute.

QQ Plots

For inter-arrival times less than approximately 2.5 seconds the Exponential fits relatively well, however, for trade inter-arrivals greater than 2.5 seconds the empirical quantiles are larger than the empirical quantiles suggesting that inter-arrival times of trades decay slower than the Exponential distribution.

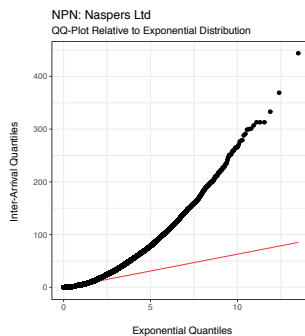
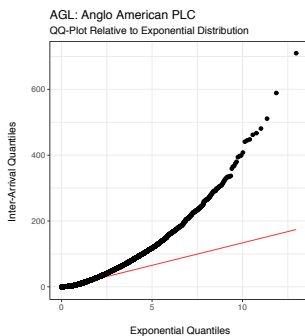


Figure: QQ Plot Relative to Exponential Distribution.

ACF

Indicates that if one observes a market order now, based on this information alone there is some non-vanishing predictability of the market order arrival times

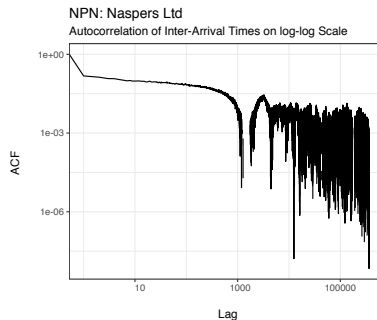
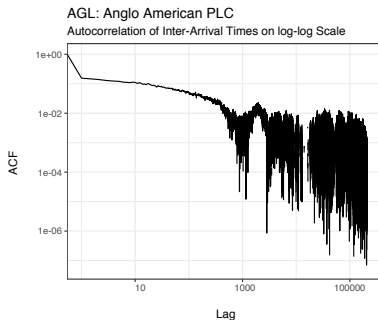


Figure: Auto-Correlation of Inter-Arrival Times.

THANKS!

References



Jean-Philippe Bouchaud, J Doyne Farmer, and Fabrizio Lillo, **How markets slowly digest changes in supply and demand**, Handbook of financial markets: dynamics and evolution, Elsevier, 2009, pp. 57–160.



Bloomberg Database, **Bloomberg professional**, Bloomberg Subscription Service, Online, 2019.



Charles MC Lee and Mark J Ready, **Inferring trade direction from intraday data**, The Journal of Finance **46** (1991), no. 2, 733–746.



Bence Toth, Imon Palit, Fabrizio Lillo, and J Doyne Farmer, **Why is equity order flow so persistent?**, Journal of Economic Dynamics and Control **51** (2015), 218–239.