ECOLE POLYTECHNIQUE DE TUNISIE

Graduation Internship Presentation

Predictive Modeling for Intraday Trading Volume

Host Organization: Quant-Dev Quantitative Analytics

February – June 2020

Elaborated by: **Youssef AMDOUNI**

Graduate Engineering Student

Supervised by : **Mr. Fathi ELACHHAB (EPT)**

Mr. Wajdi TEKAYA (Quant-Dev)

Mr. Oussema GAZZEH (Quant-Dev)

Academic Year 2019-2020

Quant-Dev
Quantitative Analytics

INTRODUCTION

- Advance in computer technology and trading instrument.

  → Trade from everywhere in the world with an immediacy of execution.

- Large variety of advanced trading algorithms.

  → Improve their trading strategies.

Effective Market Modeling & Accurate Forecast

- Implement a model to forecast the coming short-term trading volume in the Chinese Stock Market.

  → Turnover instead of volume to make data comparable.

  → Decomposing the turnover into Common Component and Specific Component.

  → A rolling mean approach will be used to predict the common component and a simple linear regression model to forecast the specific component.

CONTENTS

1 - GENERAL CONCEPT OF THE PROJECT

- HOST COMPANY PRESENTATION
- OVERVIEW OF TRADING AND MARKETS
- PROBLEM STATEMENT AND PROJECT GOALS

GENERAL CONCEPT OF THE PROJECT

HOST COMPANY PRESENTATION: QUANT-DEV

Overview:

- Consulting Company
- Established in 2014
- Headquartered in Tunis, Tunisia.

Expertise:

- High Frequency Trading
- Quantitative and Algorithmic Trading
- Data Science

GENERAL CONCEPT OF THE PROJECT

OVERVIEW OF TRADING AND MARKETS

Modern Markets

- More affordable and powerful trading platforms.
- Trade from different part of the world.
- Availability of high-frequency data.

Algorithm Trading

- Immediacy and precision of the transaction execution.
- Advanced machine learning models to support investors strategies.

High-Frequency Trading

- Common term referring to a form of algorithmic trading.
- Multiple markets evaluations and high-speed order execution.
- Fully automated trading systems.

GENERAL CONCEPT OF THE PROJECT

PROBLEM STATEMENT AND PROJECT GOALS

The trading cost is as important as the decision to trade because it is one of the key determinants of net returns.

The trading cost has three essential components:

- Fees and commissions: known ahead of trading (3%-5% of the volume traded).
- Opportunity cost: the cost associated with the inability to complete an order.
- Execution cost: the adverse price change following an order.

✓ Powerful Computers
✓ Advanced algorithms
✓ Exhaustive Databases

→ Rational trading decisions

“ Traders require advanced models to reduce their market impact”

- Develop a predictive model.

→ Predict the volume in the coming short-term period

2 - TECHNICAL TOOLS AND MATHEMATICAL BACKGROUND

- Principal Component Analysis (PCA)
- Regression Models
- Performance Metrics

TECHNICAL TOOLS AND MATHEMATICAL BACKGROUND

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis:

- ✓ Unsupervised machine learning technique.
- ✓ Mainly used for dimensionality reduction.

PCA approach:

- Combine the input variables in a specific way that the rotating features are statistically uncorrelated.
- Select only a subset of the new features according to how important they are for explaining the data.

TECHNICAL TOOLS AND MATHEMATICAL BACKGROUND

REGRESSION MODELS

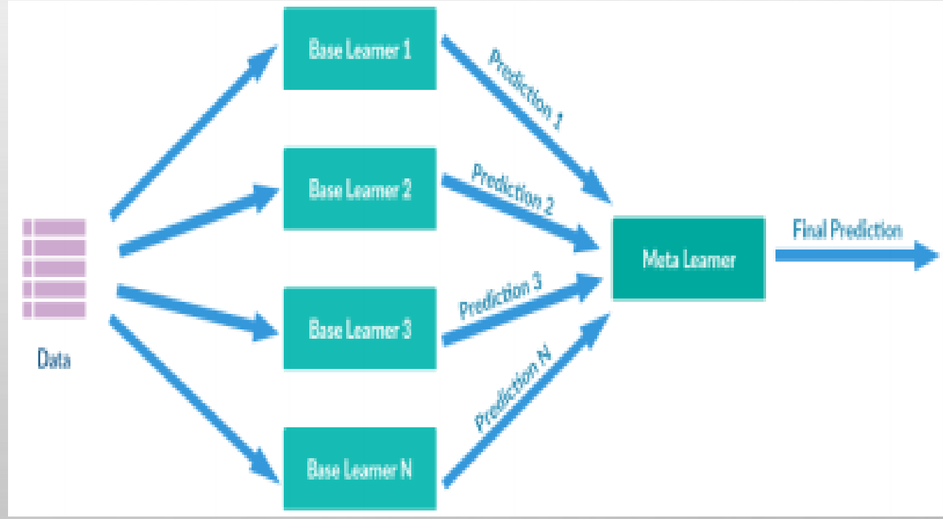
- Simple linear regression model → $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$
- Multiple linear regression model → $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_n x_n$

- Huber regressor
- Random sample consensus regressor
- Linear support vector machine

→ Regression model to deal with outliers

- Stacking regression

Uses the power of several estimators to predict the final value



TECHNICAL TOOLS AND MATHEMATICAL BACKGROUND

PERFORMANCE METRICS

- Mean Absolute Percentage Error:

$$MAPE\ (5min) = \left|\frac{predicted(5min) - actual(5\ min)}{actual\ (5min)}\right| * 100$$

$$MAPE(day) = \sum_{i=1}^{N} \frac{MAPE(5\ min)}{N}$$

$$MAPE(total) = \sum_{i=1}^{L} \frac{MAPE(day)}{L}$$

- Mean Squared Error:

$$MSE(5\ min) = (Actual\ - Predicted)^2$$

$$MSE(day) = \sum_{i=1}^{N} \frac{MSE(5\ min)}{N}$$

$$MSE(total) = \sum_{i=1}^{L} \frac{MSE(day)}{L}$$

- R Squared Score:

$$R^2 = 1 - \frac{Sum\ Squareed\ Regression\ Error\ (SSR)}{Sum\ Squared\ Total\ Error\ \ (SST)} = 1 - \frac{\sum(\ Actual_i - \widehat{Predicted_i}\ )^2}{\sum(\ Actual_i - \overline{Actual_i}\ )^2}$$

3 - EMPIRICAL ANALYSIS AND MODELING

- General Description
- Data exploration
- Modeling
- Evaluation Summary

MODELING AND EXPERIMENTAL RESULTS

GENERAL DESCRIPTION

Project goals Recall

Analyse the volume in the Chinese Stock Market by implementing a predictive model to forecast the coming short-term volume.

Decompose the volume into:

- Common component ⟺ Volume changes due to market evolution
  - Forecast → Rolling average of the previous days
  - Forecast → Tuning portfolio composition
- Specific component ⟺ Stock specific volume pattern
  - Forecast → Linear regression model
  - Forecast → Advanced regression model

Data Description

Minutely Data set from the shanghai stock exchange

- Data for 32 days (train data is three day rolling)
  → From 2020-03-05 to 2020-04-20
- Data for 50 symbols
  → Most liquid stocks.

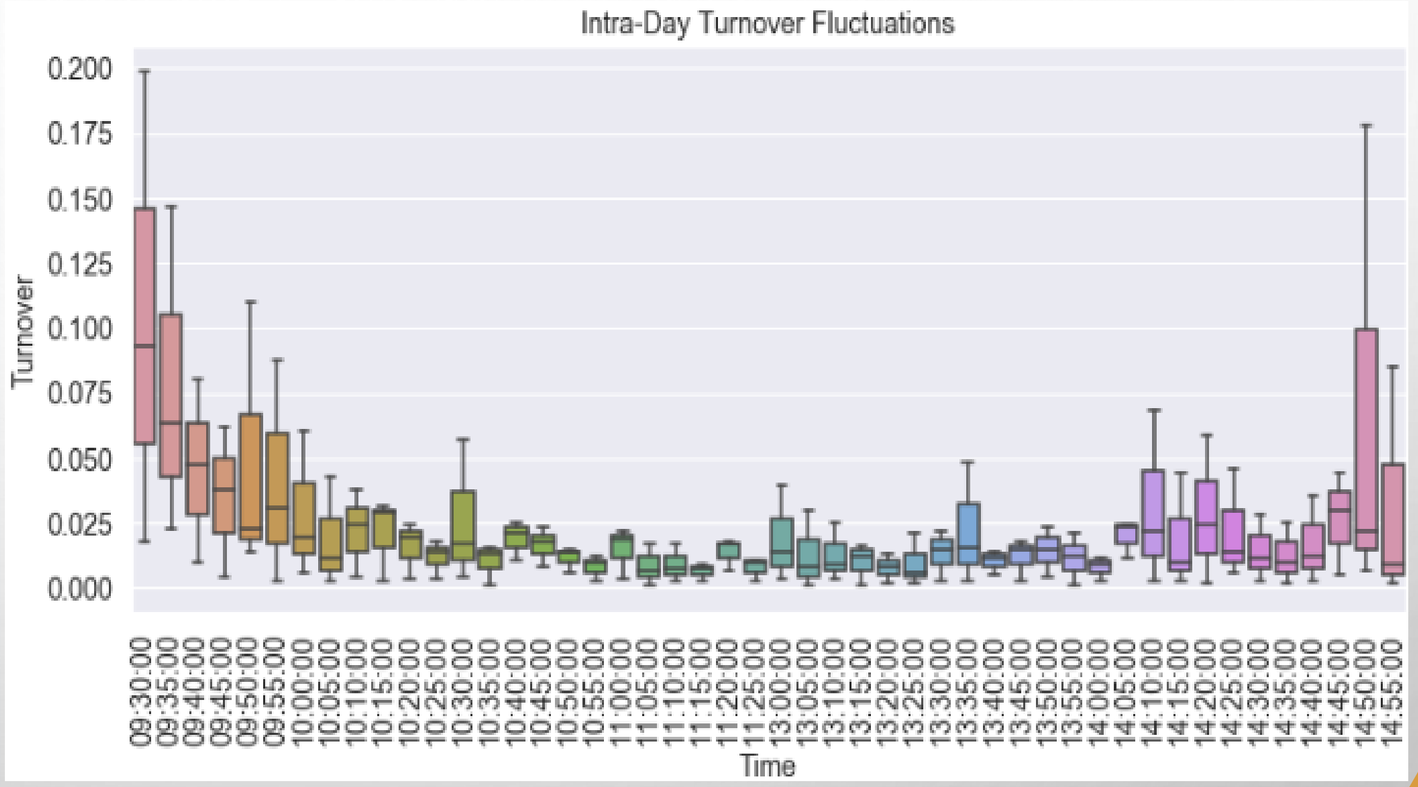
MODELING AND EXPERIMENTAL RESULTS

DATA EXPLORATION

Data Preparation:

- We split the trading day into N intervals of 5 minutes.
- We use the turnover to make the data of all stocks comparable

$$Turnover_{m,t,i} = \frac{Shares\ Traded_{m,t,i}}{Outstanding\ Shares}$$



The turnover plot has an U-shaped distribution

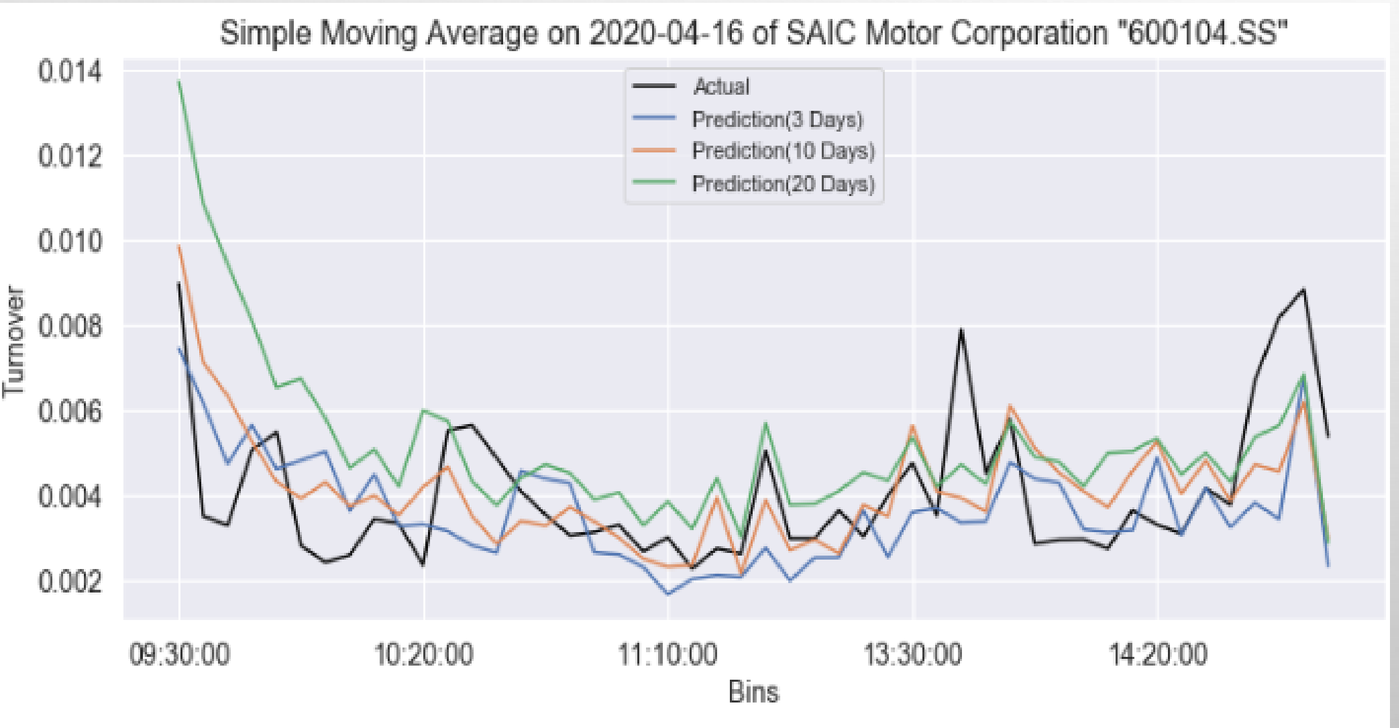
MODELING AND EXPERIMENTAL RESULTS

MODELING(1/5)

Benchmark : Simple Moving Average Strategy (SMA)

Assumption:

- The turnover distribution is approximately the same every trading day.
- The intra-day turnover has a U-shape distribution.



Prediction:

$$x_{N+1,i,m} = \sum_{t=1}^{N} \frac{x_{t,i,m}}{N}$$

$x_{t,i,m}$: Turnover of stock m in bin I on day t.

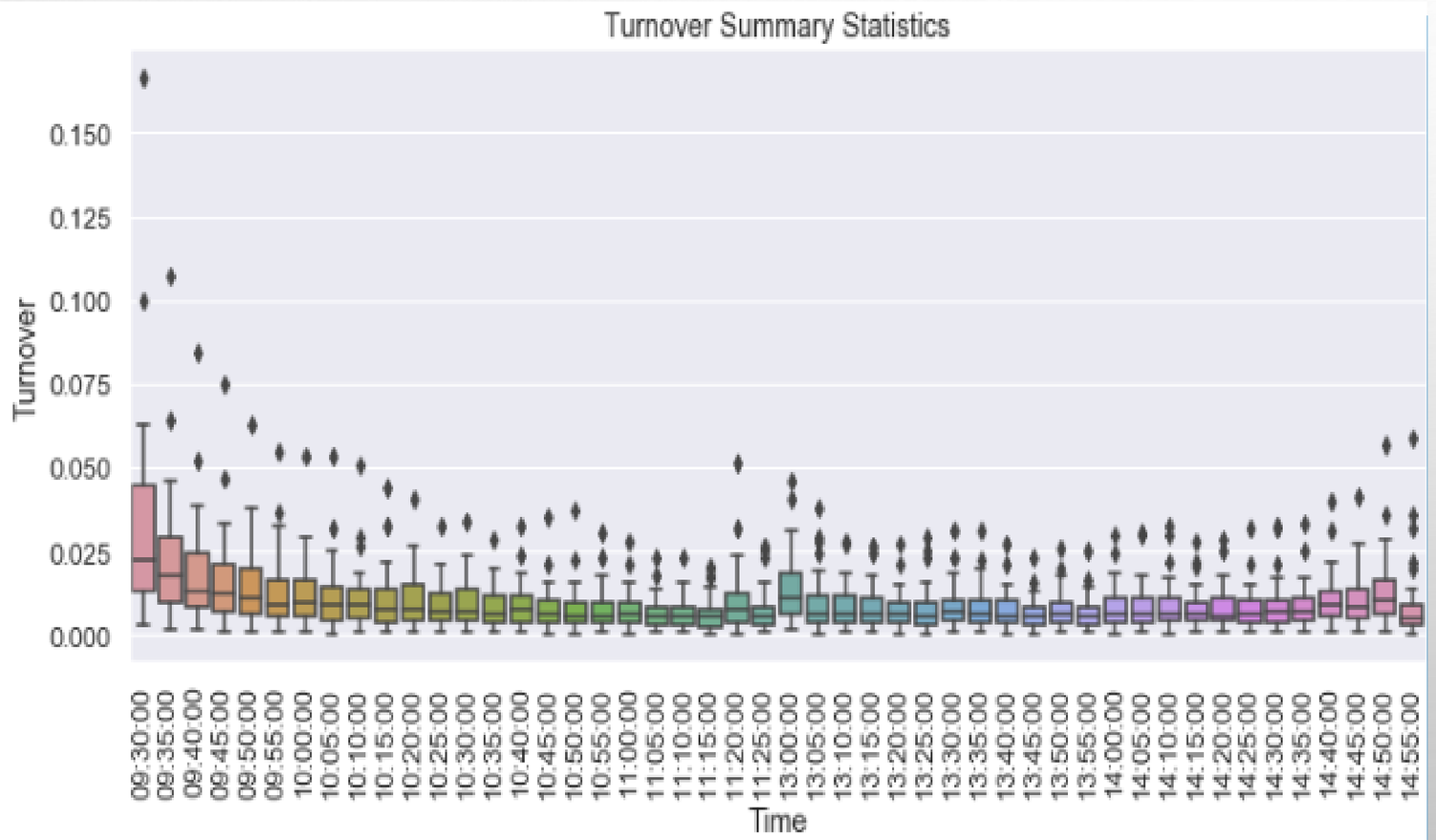
N : Time window (three days).

MODELING AND EXPERIMENTAL RESULTS

MODELING(2/5)

SMA Limitations

Summary statistics of the intra day turnover data of 50 symbols:



- ✓ Q95 is approximately 13 times larger than the Q5 and almost 3 times larger than the mean.
- ✓ The mean only cannot be well representative of turnover variation.
- ✓ There is some random fluctuations.

→ A static method does not warn us about the updated turnover information.

MODELING AND EXPERIMENTAL RESULTS

MODELING(3/5)

Dynamic Models

Additive Model

- Common Component: Turnover changes due to market evolution.

  Forecast → Rolling average of previous days

- Specific Component: Stock's specific turnover pattern

  Forecast →
  - ✓ Autoregressive Moving average (1, 1)
  - ✓ Support Vector Regressor
  - ✓ Partial Least Squares

MODELING AND EXPERIMENTAL RESULTS

MODELING(4/5)

Dynamic Strategy '2-Component Model'

Assumption:

The turnover has two component:

- Common Component: Turnover changes due to market evolution.
- Specific Component: Stock's specific turnover pattern

$$X_{t,i,m} = C_{t,i,m} + R_{t,i,m}$$

$X_{t,i,m}$: Turnover of stock m during period i on day t

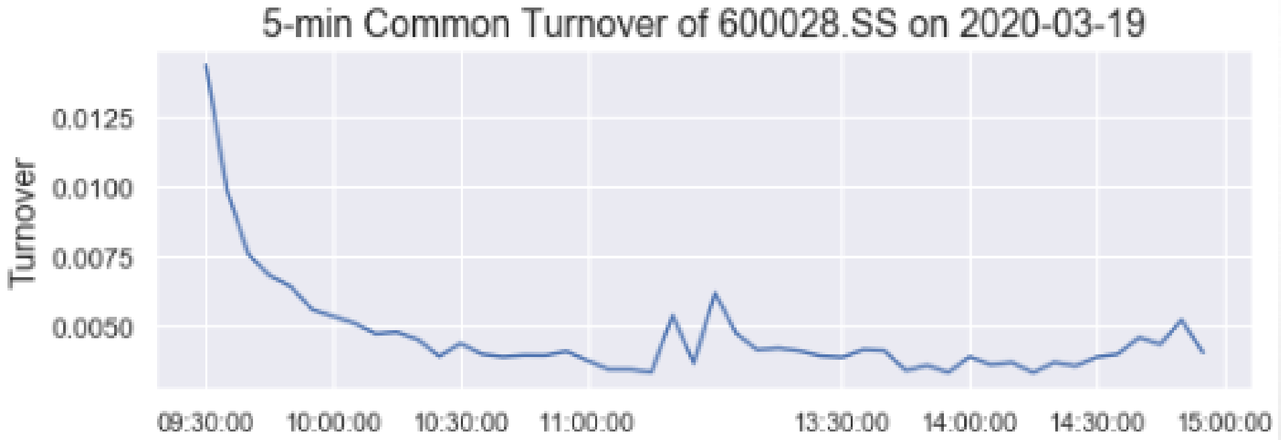
$R_{t,i,m}$: Specific turnover of stock m during period i on day t

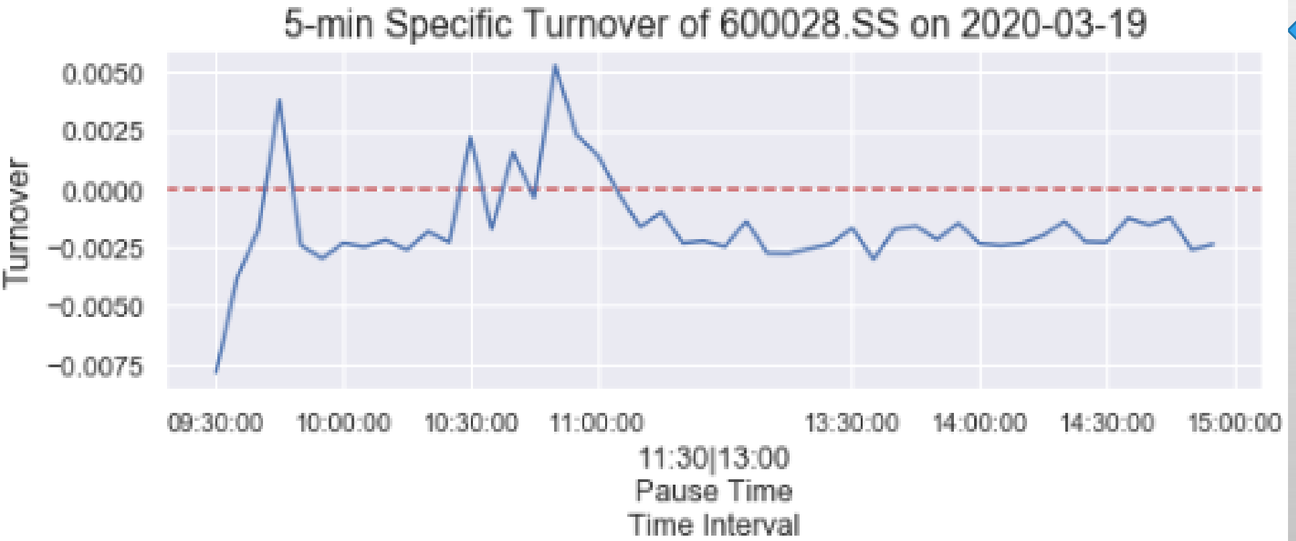
$C_{t,i,m}$: Common component turnover of stock m during period i on day t

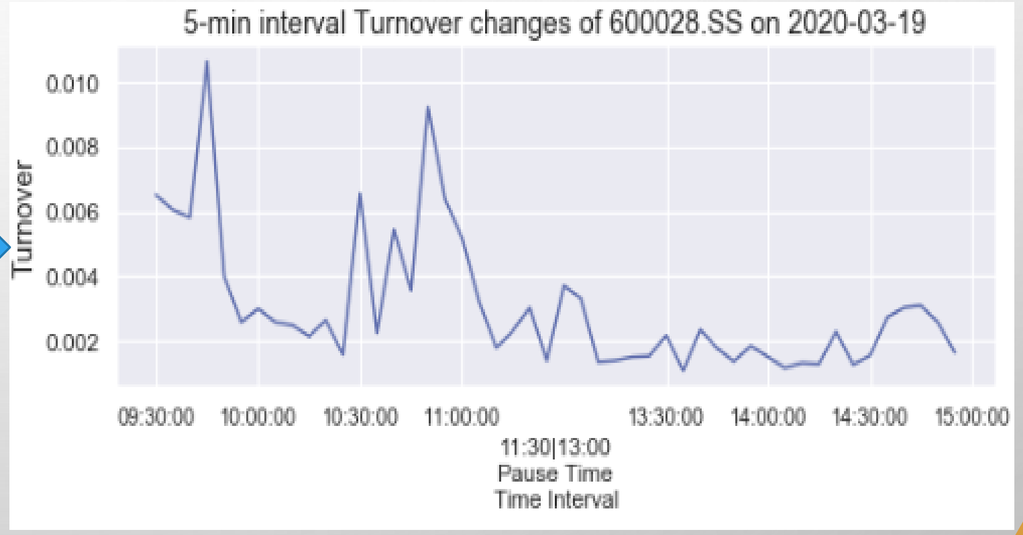
MODELING AND EXPERIMENTAL RESULTS

MODELING(5/5)

Dynamic Strategy '2-Component Model'







MODELING AND EXPERIMENTAL RESULTS

EVALUATION SUMMARY (1/5)

Common Component Improvement

	Count	MAPE Mean	MAPE Std	MSE Mean	MSE Std	R^2 Score
SMA	69600	0.833	1.80	0.0250	0.58	0.146
Common Component	69600	0.810	1.61	0.0240	0.58	0.181
2-Component Model 50	69600	0.688	1.20	0.0195	0.57	0.335
By Industry	69600	0.721	1.53	0.0229	0.59	0.218
2-Component Model 25	69600	0.693	1.31	0.0211	0.58	0.279
2-Component Model 10	69600	0.685	1.35	0.0223	0.58	0.238
2-Component Model 5	69600	0.697	1.44	0.0230	0.58	0.216
2-Component Model 2	69600	0.738	1.69	0.0231	0.58	0.212

By Industry 2-Component Model:

Constructs portfolios contain only stocks from one industry.

By Groups 2-Component Model:

Constructs portfolios using smaller groups of stocks.

MODELING AND EXPERIMENTAL RESULTS

EVALUATION SUMMARY (2/5)

Specific Component Improvement

2-Component Model Multi Order 2:

$$\hat{R}_{m,i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{m,i-1,t} + \hat{\beta}_2 R_{m,i-2,t}$$

2-Component Model Multi Order 3:

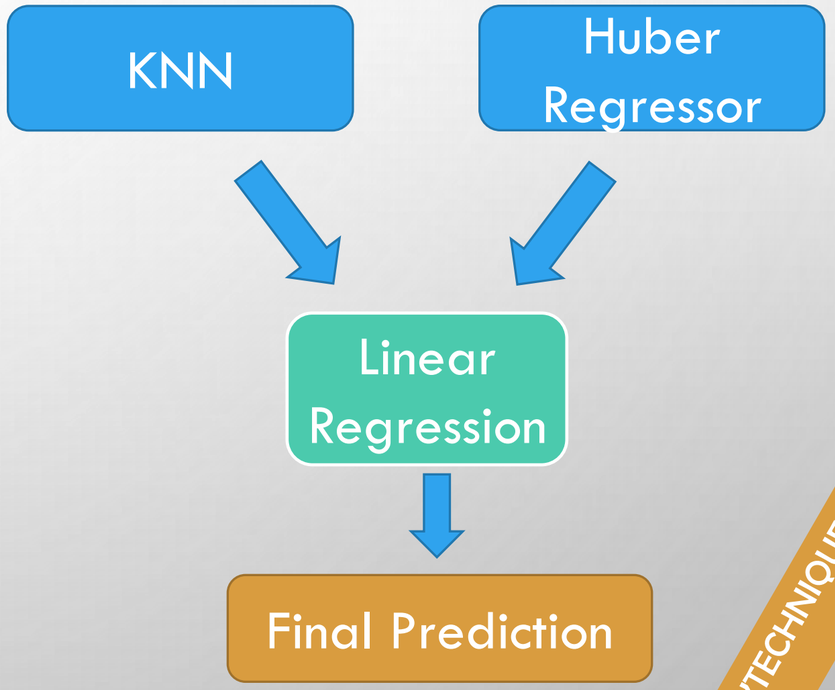
$$\hat{R}_{m,i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{m,i-1,t} + \hat{\beta}_2 R_{m,i-2,t} + \hat{\beta}_3 R_{m,i-3,t}$$

2-Component Huber
2-Component LinearSVR
} Penalizes abnormal fluctuation

2-Component RANSAC → Selects a subset of inlier to fit the model

2-Component KNN → Predicts the new observation by averaging the closets training data

2-Component Model Stacking



MODELING AND EXPERIMENTAL RESULTS

EVALUATION SUMMARY (3/5)

Specific Component Improvement

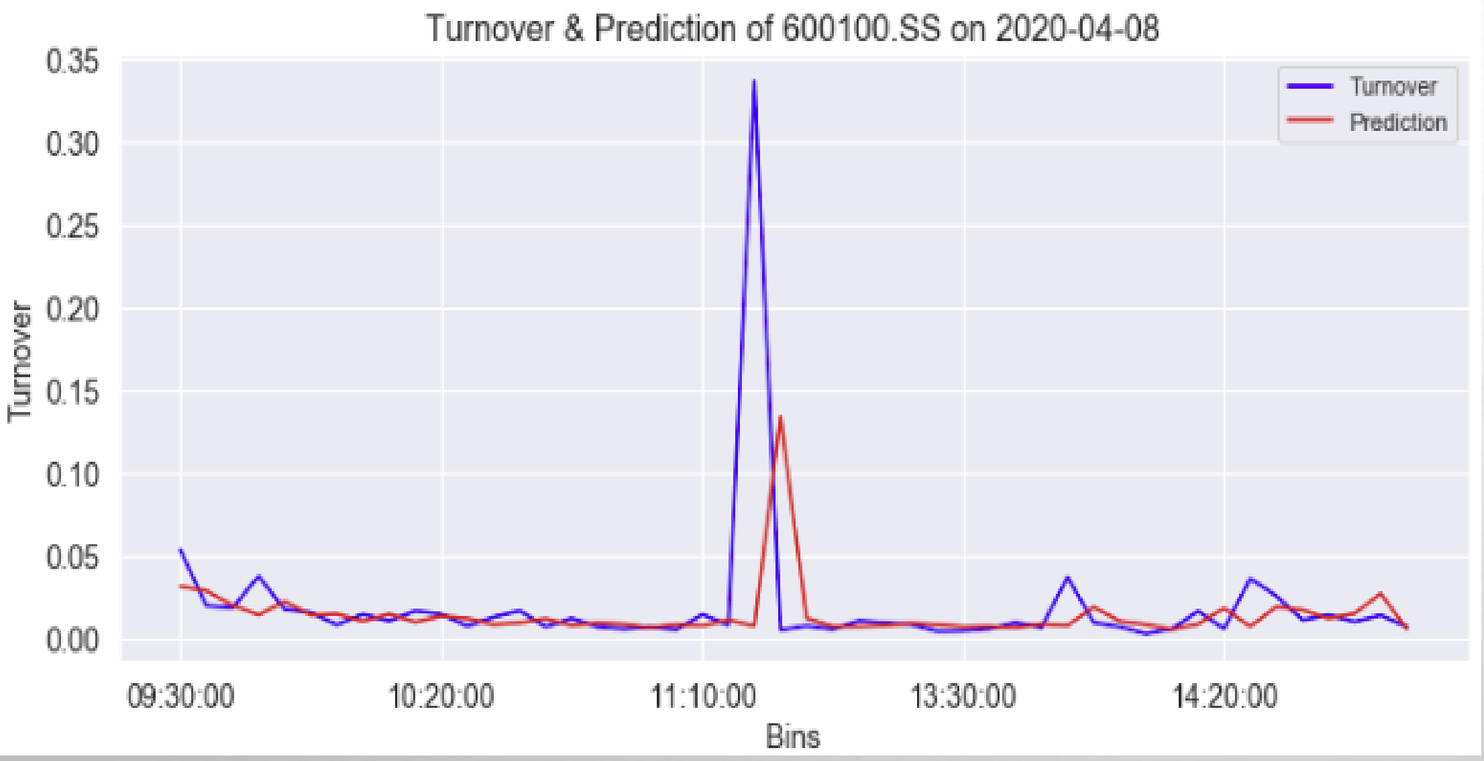
	Count	MAPE Mean	MAPE Std	MSE Mean	MSE Std	R^2 Score
Common Component	69600	0.810	1.61	0.0240	0.58	0.181
2-Component Model*	69600	0.688	1.20	0.0195	0.57	0.335
2-Component Model Multi Order 2	69600	0.671	1.11	0.0200	0.57	0.317
2-Component Model Multi Order 3	69600	0.667	1.07	0.0203	0.57	0.306
2-Component Model HuberRegressor	69600	0.598	1.17	0.0196	0.58	0.331
2-Component Model LinearSVR	69600	0.646	1.24	0.0196	0.58	0.330
2-Component Model KNN	69600	0.687	1.20	0.0191	0.57	0.348
2-Component Model RANSACRegressor	69600	0.590	1.26	0.0206	0.60	0.299
2-Component Model Stacking	69600	0.672	1.26	0.0191	0.56	0.348

Advanced Machine learning models can improve the model performance

MODELING AND EXPERIMENTAL RESULTS

EVALUATION SUMMARY (4/5)

2-Component Model Limitations



Fails to forecast turnover shocks

MODELING AND EXPERIMENTAL RESULTS

EVALUATION SUMMARY (5/5)

Improvement Summary:

	Count	MAPE Mean	MAPE Std	MSE Mean	MSE Std	R^2 Score
SMA	69600	0.833	1.80	0.0250	0.58	0.146
2-Component Model*	69600	0.688	1.20	0.0195	0.57	0.335
2-Component Model HuberRegressor	69600	0.598	1.17	0.0196	0.58	0.331

CONCLUSION AND PERSPECTIVES

- Reduce market impact:

  → A model to predict the coming five minutes turnover.

  Decompose the turnover into a common component and specific component:
  - Rolling mean approach to predict the common component.
  - Regression model to predict the actual bin specific component.
  - Common and specific components improvement.

- Future prospects :

  ✓ Think about other features to forecast the specific component.

  ✓ Think about multiplicative model.
  - A daily component
  - An intraday periodic component
  - An intraday non-periodic

"DATA IS WHAT YOU NEED TO DO

ANALYTICS.

INFORMATION IS WHAT YOU NEED TO DO

BUSINESS."

John Owen

Thanks

ECOLE POLYTECHNIQUE DE TUNISIE

Graduation Internship Presentation

Predictive Modeling for Intraday Trading Volume

Host Organization: Quant-Dev Quantitative Analytics

February – June 2020

Elaborated by: **Youssef AMDOUNI**

Graduate Engineering Student

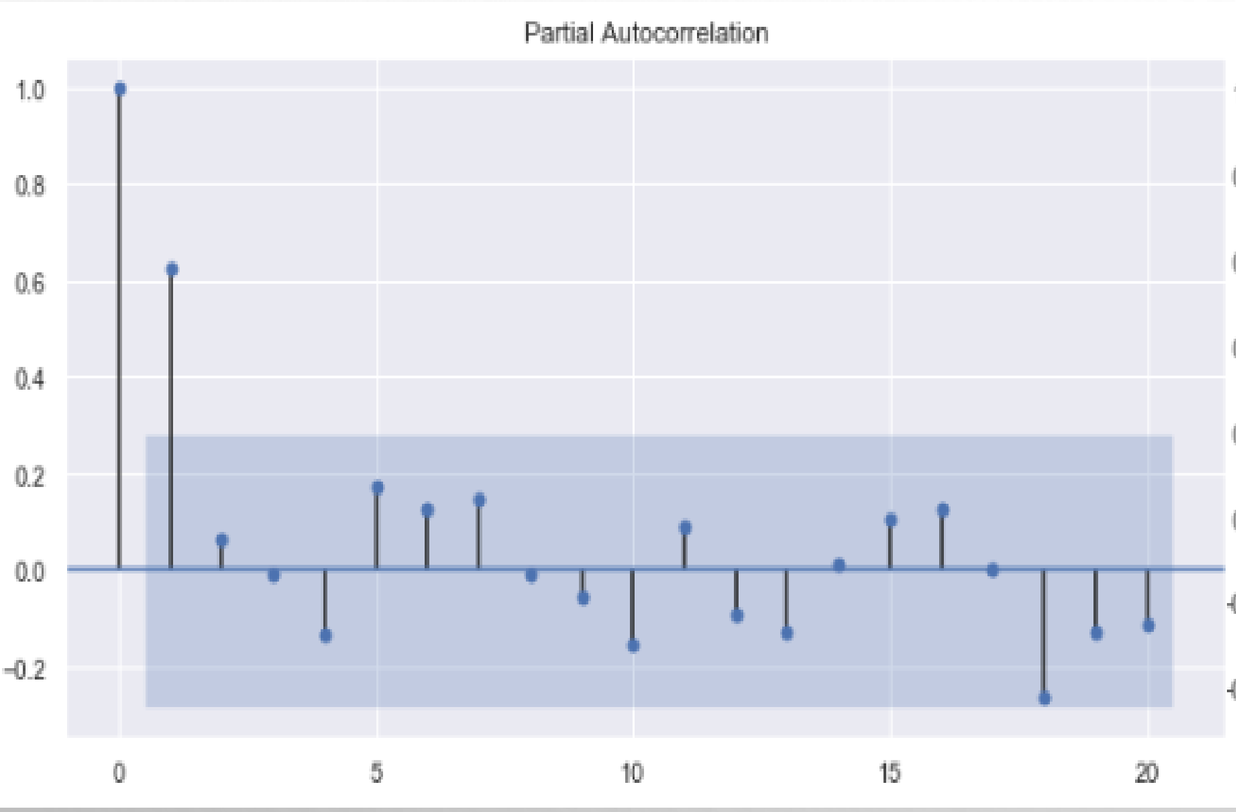
Supervised by : **Mr. Wajdi TEKAYA**

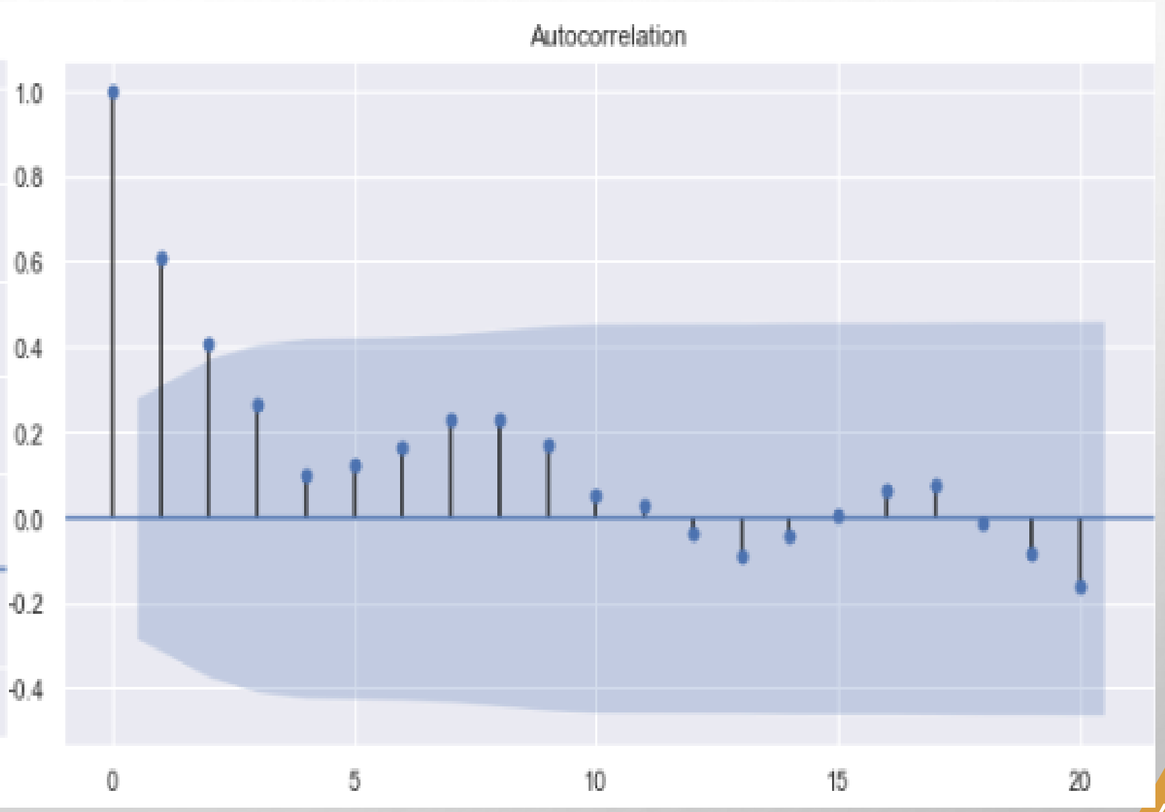
Mr. Oussema GAZZEH

Mr. Fathi ELACHHAB

Academic Year 2019-2020

APPENDICES


Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0
5
10
15
20

APPENDICES

Linearity