MS&E 448

Trading forex with distributed limit order book

Jingbo Yang, Jon Braatz, Xiaoye Yuan, Carolyn Kao, Jiachen Ge, Sunny Shah

With data provided by Integral
Under guidance of Dr. Lisa Borland and Enguerrand Horel.

Outline

1. Intro/Problem Statement
2. Data
    a. What does each data point look like
    b. Statistics
        i. Across time
        ii. Across currency pairs (e.g. correlations)
        iii. Across LPs (I haven't thought of a good one for this yet)
3. Related Work / Existing Methods
4. Methods
    a. Baseline method
    b. Next steps method

Data from Integral

8 Currency pairs

- USDCAD, USDCHF, USDJPY, USDSEK,
- AUDUSD, EURUSD, GBPUSD, NZDUSD

Across 1 month, 5 LPs

- February 1st, 2019 - March 1st, 2019
- Sunday: starting at 1800 (discard, too few trades)
- Monday-Thursday: 24 hours
- Friday: Ends at 2200 (discard after 1800)
- 25 days, ~400 active hours

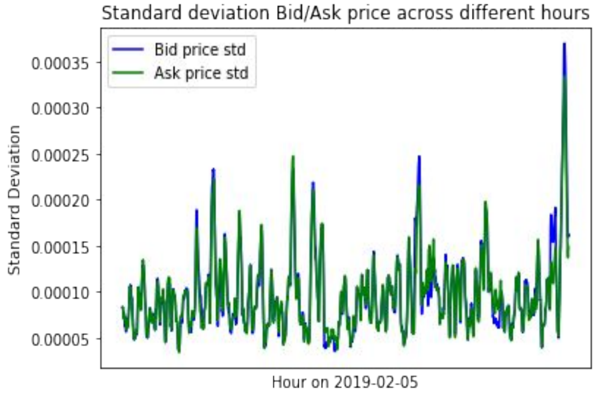
Data from Integral

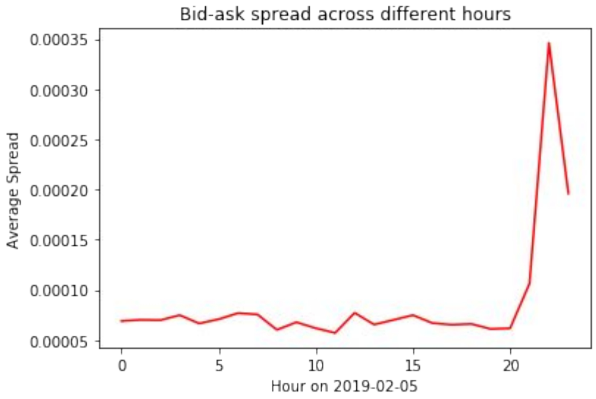
provider	currency pair	time	bid price	bid volume	ask price	ask volume
LP-1	EURUSD	02.25.2019 00:00:00.819	1.13417	1000000	1.13424	1000000
LP-1	EURUSD	02.25.2019 00:00:00.819	1.13417	1000000	1.13423	1000000
LP-1	EURUSD	02.25.2019 00:00:00.819	1.13417	1000000	1.13423	1000000
LP-1	EURUSD	02.25.2019 00:00:00.841	1.13411	1000000	1.13423	1000000
LP-1	EURUSD	02.25.2019 00:00:00.841	1.13411	1000000	1.13423	1000000
LP-1	EURUSD	02.25.2019 00:00:00.841	1.13411	1000000	1.13423	1000000
LP-1	EURUSD	02.25.2019 00:00:00.895	1.1341	1000000	1.13422	1000000
LP-1	EURUSD	02.25.2019 00:00:00.896	1.1341	1000000	1.13422	1000000
LP-1	EURUSD	02.25.2019 00:00:00.896	1.1341	1000000	1.13422	1000000
LP-1	EURUSD	02.25.2019 00:00:00.940	1.13414	1000000	1.13421	1000000
LP-1	EURUSD	02.25.2019 00:00:00.940	1.13414	1000000	1.13421	1000000
LP-1	EURUSD	02.25.2019 00:00:00.940	1.13414	1000000	1.13421	1000000
LP-1	EURUSD	02.25.2019 00:00:00.958	1.13414	1000000	1.1342	1000000
LP-1	EURUSD	02.25.2019 00:00:00.958	1.13414	1000000	1.1342	1000000
LP-1	EURUSD	02.25.2019 00:00:00.959	1.13414	1000000	1.1342	1000000
LP-1	EURUSD	02.25.2019 00:00:01.039	1.13414	1000000	1.13421	1000000
LP-1	EURUSD	02.25.2019 00:00:01.039	1.13414	1000000	1.13421	1000000
LP-1	EURUSD	02.25.2019 00:00:01.039	1.13414	1000000	1.13421	1000000
LP-1	EURUSD	02.25.2019 00:00:01.671	1.13412	1000000	1.13419	1000000
LP-1	EURUSD	02.25.2019 00:00:01.671	1.13412	1000000	1.13419	1000000
LP-1	EURUSD	02.25.2019 00:00:01.671	1.13412	1000000	1.13419	1000000
LP-1	EURUSD	02.25.2019 00:00:01.734	1.13408	1000000	1.1342	1000000
LP-1	EURUSD	02.25.2019 00:00:01.734	1.13408	1000000	1.1342	1000000

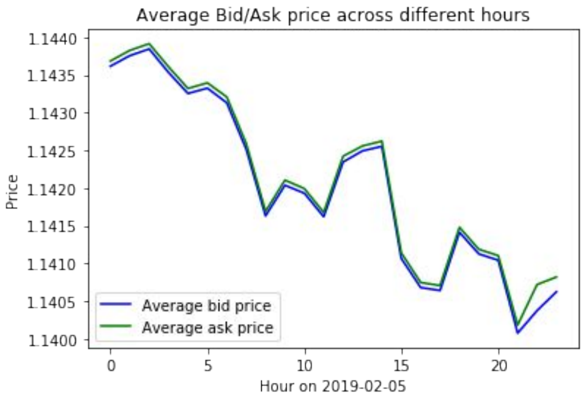
- Liquidity Provider
- Currency pair
- Exact Time
- Bid price
- Bid volume
- Ask price
- Ask volume

Example Statistics

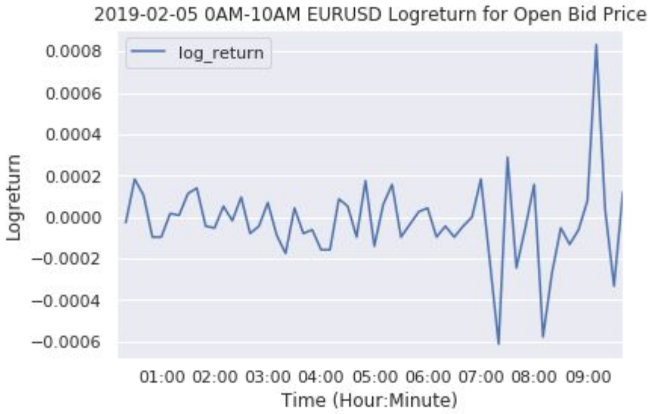
EUR/USD Feb. 5 th, 2019

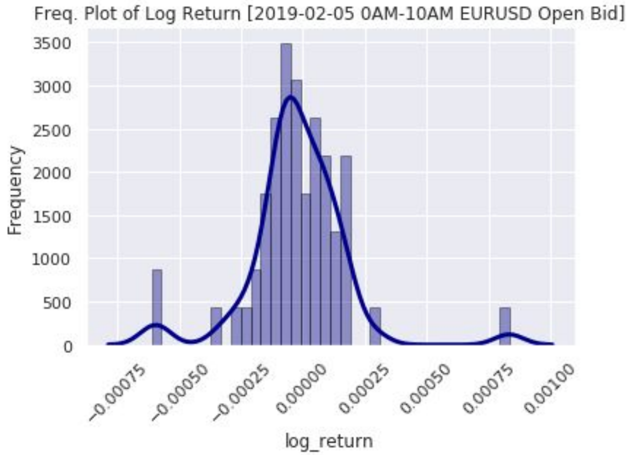


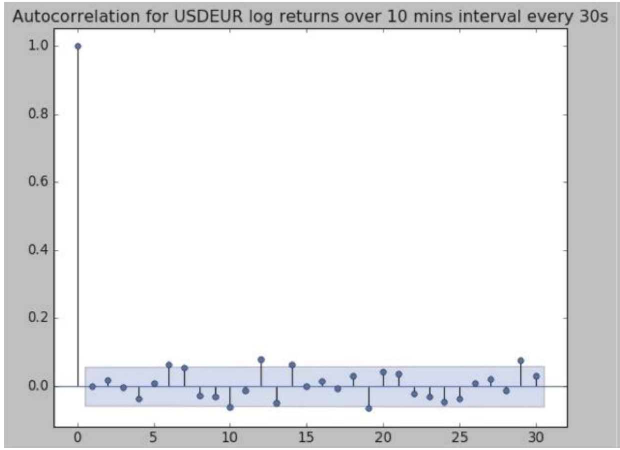


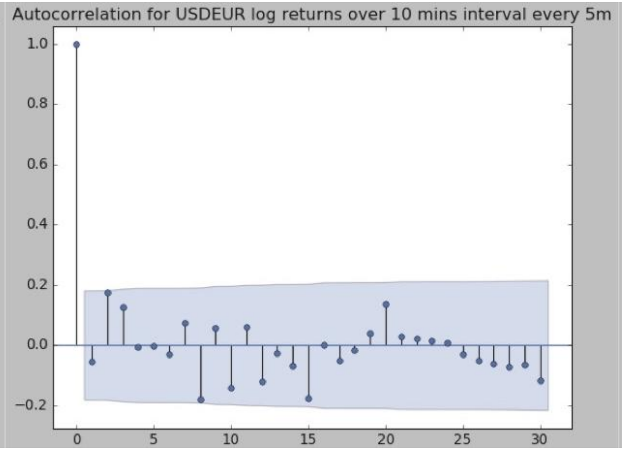


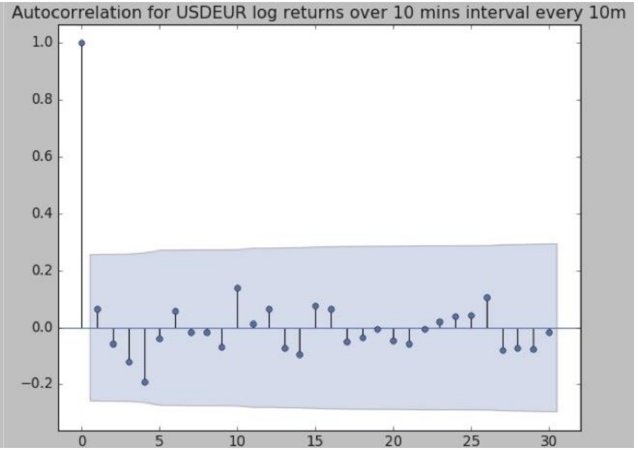
Example Statistics



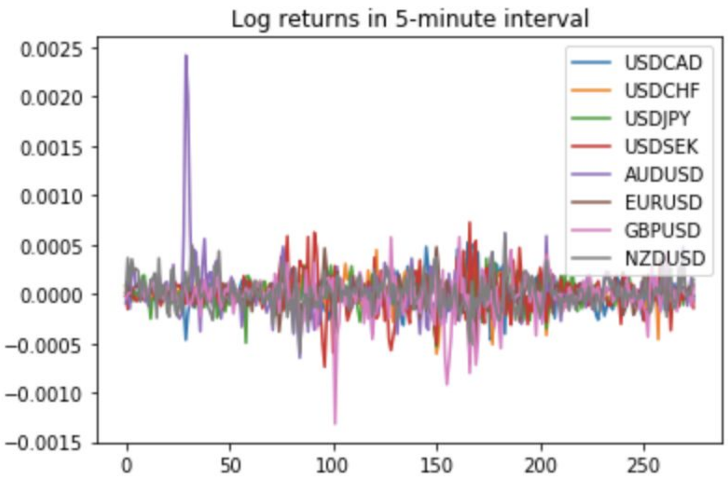


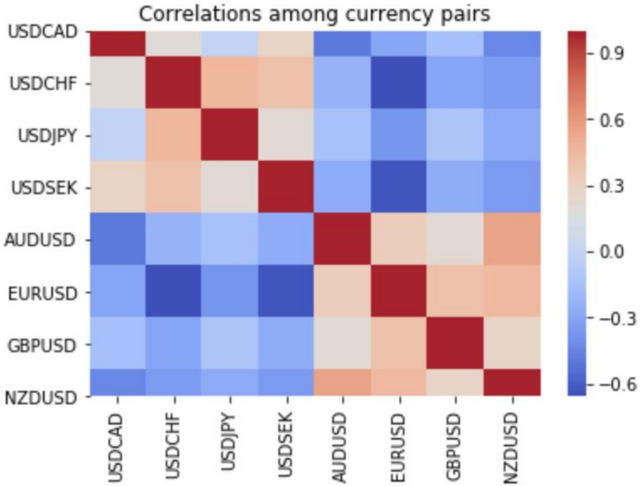


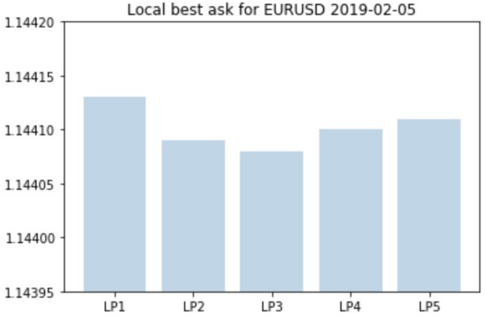


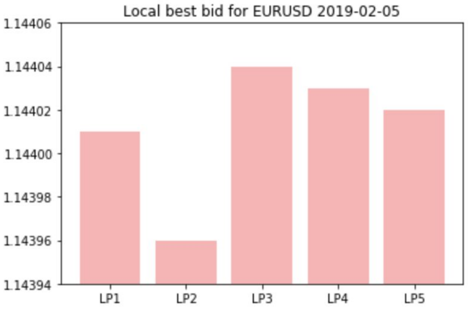


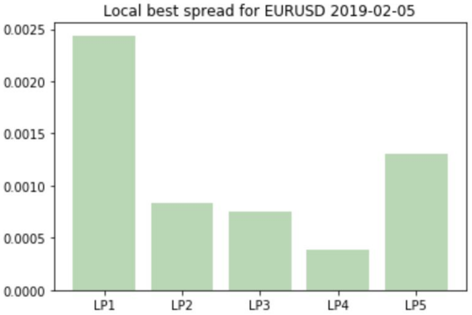
Example Statistics







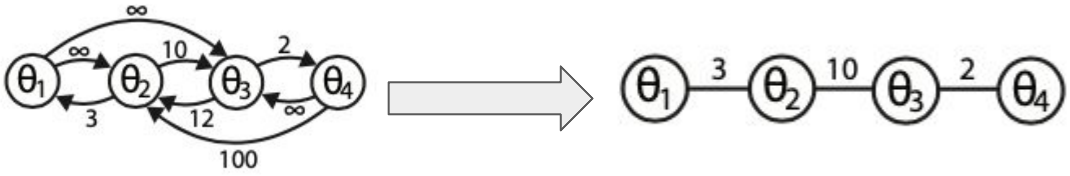




Related Works

Quasi-centralized limit order books (QCLOBs) (Gould, Porter, and Howison, 2017)

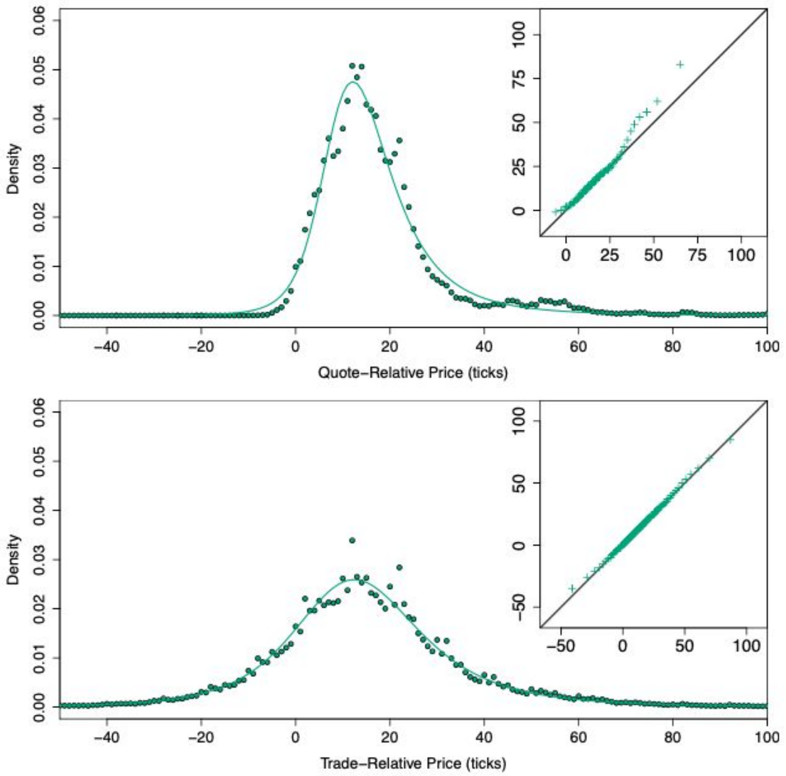
- Financial institutions access only the trading opportunities offered by counterparties with whom they possess sufficient bilateral credit



- QCLOB is global but LOB for each LP is local
  - QCLOB is not visible to all LPs

Statistics of QCLOBs

- Obtained data from the Hotspot FX platform
- Results
  - For all queue lengths, the mean size of arriving marketing orders is *strictly* smaller than the queue length
  - The authors use the generalized *t*-distribution to model the distribution of the limit orders for EUR/USD in one day



Reinforcement Learning for FX Trading (2019 Group): Policy-Based RL

- Picked a currency pair to trade at best bids and offers across LPs
- States
  - Previous action and best bids and asks of all currencies over last 8 time steps
- Actions
  - How much to short/long the pair (as a fraction of current quote currency holdings)
- Model
  - 3-layer policy network trained using REINFORCE (Monte Carlo policy gradient), with dropout and SGD

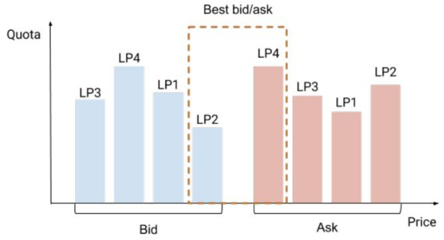
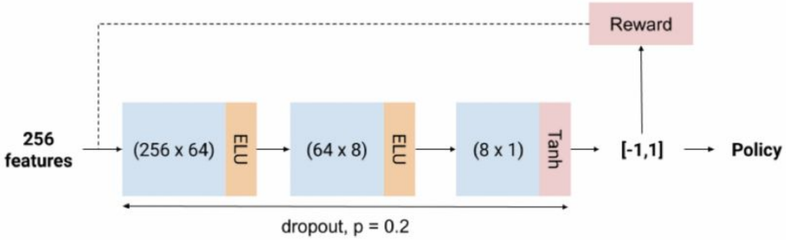


Figure 2: Order Book



Algorithm 1: Deep reinforcement learning

Initialize: Differentiable policy parameterization $\pi(a|s,\theta)$ (i.e., trading agent)
for $l \leftarrow 0$ **to** L **do**
  Generate a new episode $(s_0, a_0, r_1, ..., a_{\tau-1}, r_\tau)$ following current $\pi(a|s,\theta)$
  for $t \leftarrow 0$ **to** $\tau = 3,600$ **do**
    Cumulative return $G \leftarrow$ return from step t (G_t)
    $\theta \leftarrow \theta + \alpha\gamma^t G \nabla_\theta \ln \pi(a_t|s_t,\theta)$
    $t = t + 1$
  end
end

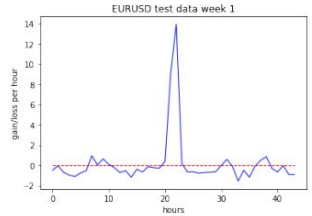
Reinforcement Learning for FX Trading (2019 Group): Policy-Based RL

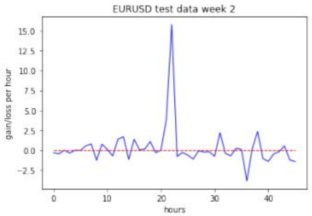
	2/1	2/3	2/4	2/5	2/6	Train Week 1
Eval/test Week 1	2/7	2/8	2/10	2/11	2/12	Train Week 2
Eval/test Week 2	2/13	2/14	2/15	2/17	2/18	Train Week 3
Eval/test Week 3	2/19	2/20	2/21	2/22	2/24	
	2/25	2/26	2/27	2/28	3/1	

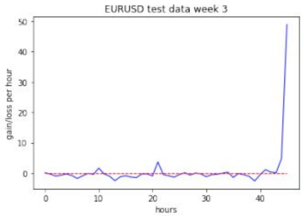
- Experiments
  - Trained on 3 different currency pairs
  - Maximized profit over hour long episodes chosen from the training window
  - for three different weeks, using the first four days of the week as the training set, the fifth day as the eval/val set, and the sixth day as the test set.
- Results: inconclusive

Table 3: EURUSD model performance summary on test data set

	week 1	week 2	week 3
Mean Return ($)	0.18	0.30	0.69
Variance ($)	2.758	2.593	7.291
Yield	0.014%	0.023%	0.052%







Baseline Model

- Input
  - Bid, ask and volume from across LP for 1 currency pair
  - Lookback 500 quotes
- Output
  - Regression (price/fluctuation prediction)
  - Classification (trend determination, up, down, flat)
- Model
  - Gradient boosting decision tree (LightGBM)

	Price Regression (mean abs error)	Trend Classification (-1.5, +1.5, 7 classes) (accuracy)
EURUSD	32 pips	37%
AUDUSD	44 pips	26%

Current Progress

- Infrastructure setup on Google Cloud Compute
  - Cloud storage of pre-processed data
  - Shared environment for running experiments
- Data pipeline for deep learning and reinforcement learning
  - Simple API for querying

```
api = utils.DataAPI()
df_usdcad = api.get('USDCAD', '1', start_time='2019-02-01 16:30:00', end_time='2019-02-01 16:45:00')
```

  - PyTorch based data loading

Next Step

- More complex deep learning models
    - Multiple layers
    - Trend classification through embedding generation
- Adaptation for reinforcement learning models
    - Adapt/refine methods from existing RL papers
- Trading simulation environment
    - Handle multiple currency pairs
    - Properly account for bid-ask spread
    - Incorporate account information with trading model

Thanks