

# 平均回帰する資産に対する強化学習トレードモデル

2025/7/8

FE課 小林

# OU過程

- 平均回帰過程の一番簡単なモデルとして、今回はOU過程を用いる。

$$dX(t) = \theta(\mu - X(t))dt + \sigma dW(t)$$

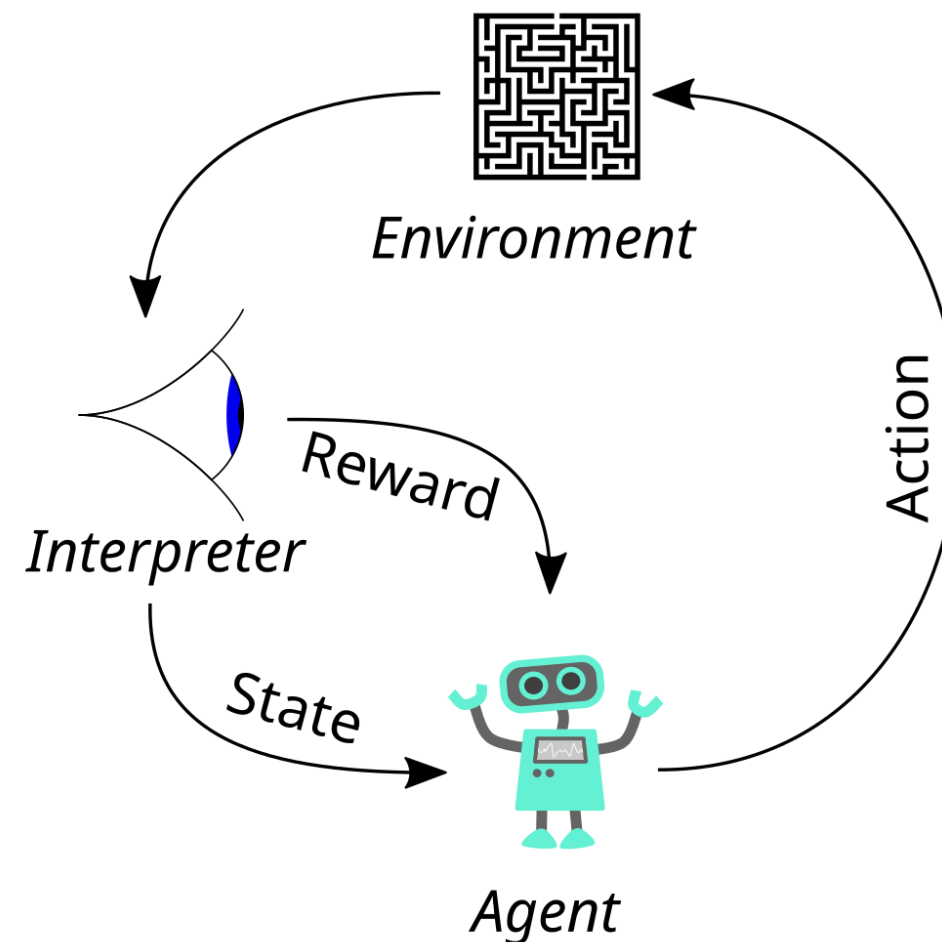
- (Wikipediaに数行で書かれてる程度の)簡単な計算で期待値や分散が求められる。

$$\begin{aligned}\mathbb{E}[X(t)] &= \mu + (X(0) - \mu)e^{-\theta t} \\ \text{Var}[X(t)] &= \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t})\end{aligned}$$

- OU過程で推移する資産に対しては、シャープレシオを最大化するトレード法の解析解が知られている。
  - 売買手数料込みで利益を最大化する方法(元論文が有料のためブログ)  
<https://hudsonthames.org/optimal-trading-thresholds-for-the-o-u-process/>
  - 利食い、損切の基準を設けた場合の最適戦略  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3534445](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3534445)

# 強化学習

- 環境とエージェントが相互作用する状態で、エージェントが最適な行動を選択できるように学習することを強化学習という。
- トレードマシンを作る場合、マーケットインパクト、トレードコスト、ポジションなどを管理した上で選択を行うため、強化学習が適している。
- 強化学習を行うアルゴリズムは複数あるが、今回は簡単なQ学習を用いた。



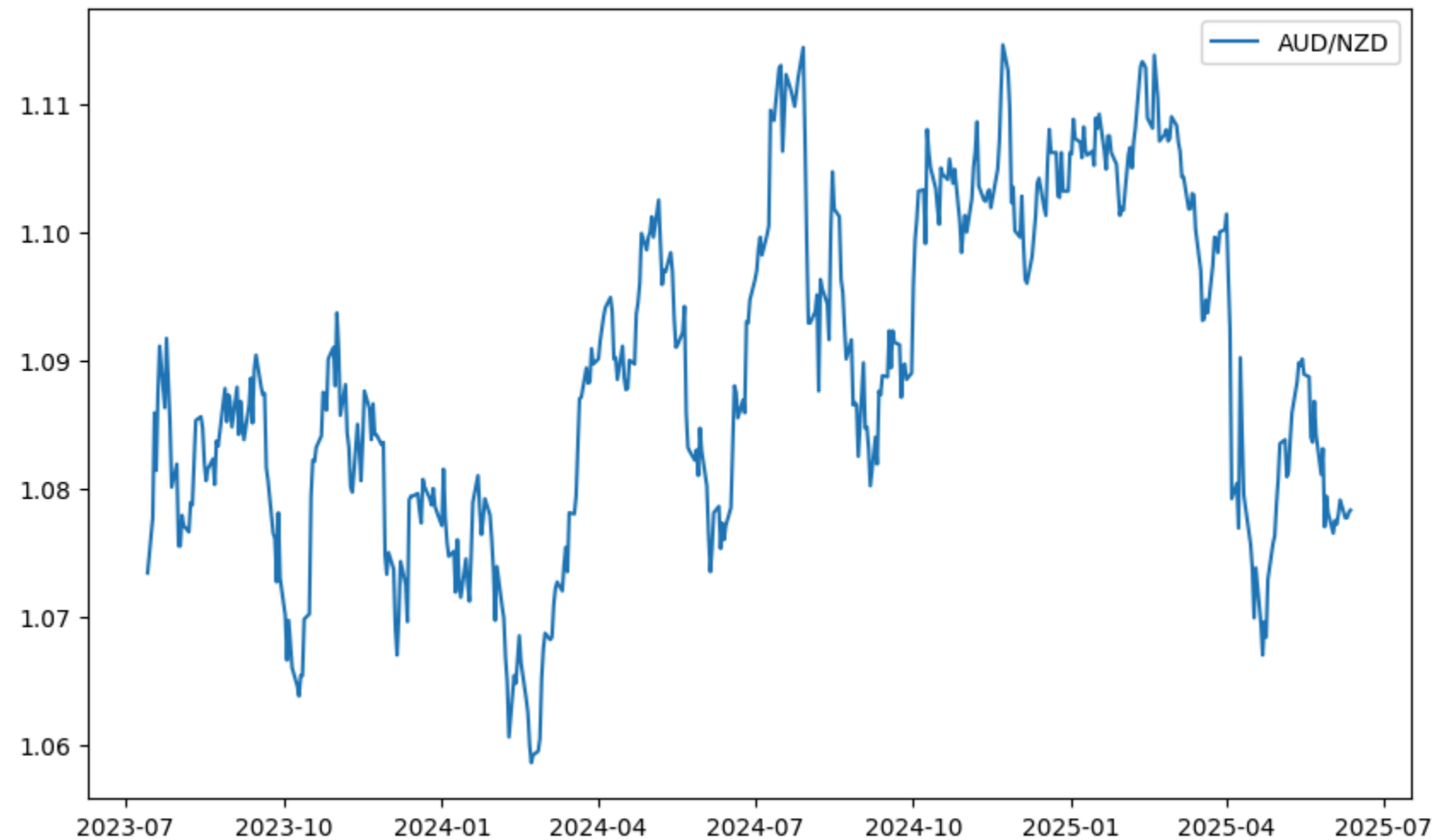
# やりたいこと

---

1. 平均回帰性を持つ通貨ペアやポジションを発見する
2. OU過程と仮定して、パラメータをキャリブレーションする
3. 同じパラメータのOU過程を生成し、学習
4. 元の実データに当てて、バックテスト

# AUD/NZDの定常性

- 2023/7/14 ~ 2025/6/12の500営業日。ADF検定でp値が0.08799。
- $(\theta, \mu, \sigma) = (0.0269, 1.0894, 0.0030)$
- そもそも定常化といわれるととても微妙だが、とりあえず進行する。



# エージェントモデル

インペリアルカレッジロンドンの教材([https://www.imperial.ac.uk/media/imperial-college/faculty-of-natural-sciences/departments-of-mathematics/math-finance/Bertermann-\\_ARVID\\_01912841.pdf](https://www.imperial.ac.uk/media/imperial-college/faculty-of-natural-sciences/departments-of-mathematics/math-finance/Bertermann-_ARVID_01912841.pdf))のp23と同じ。

**状態**  $S_t = (S_t^1, S_t^2, S_t^3) \in \mathcal{S}$

- $S_t^1 \in \mathbb{R}_{\geq 0}$ : 現在の為替
- $S_t^2 \in \mathbb{R}_{\geq 0}$ : 最後にポジションを持ち替えた時点での為替
- $S_t^3 \in \{-1, 1\}$ : 現在のポジション

**行動**  $A_t \in \mathcal{A} = \{sell, buy, hold\}$

- $S_t^3 = 1$ のとき、 $A_t \in \{sell, hold\}$
- $S_t^3 = -1$ のとき、 $A_t \in \{buy, hold\}$

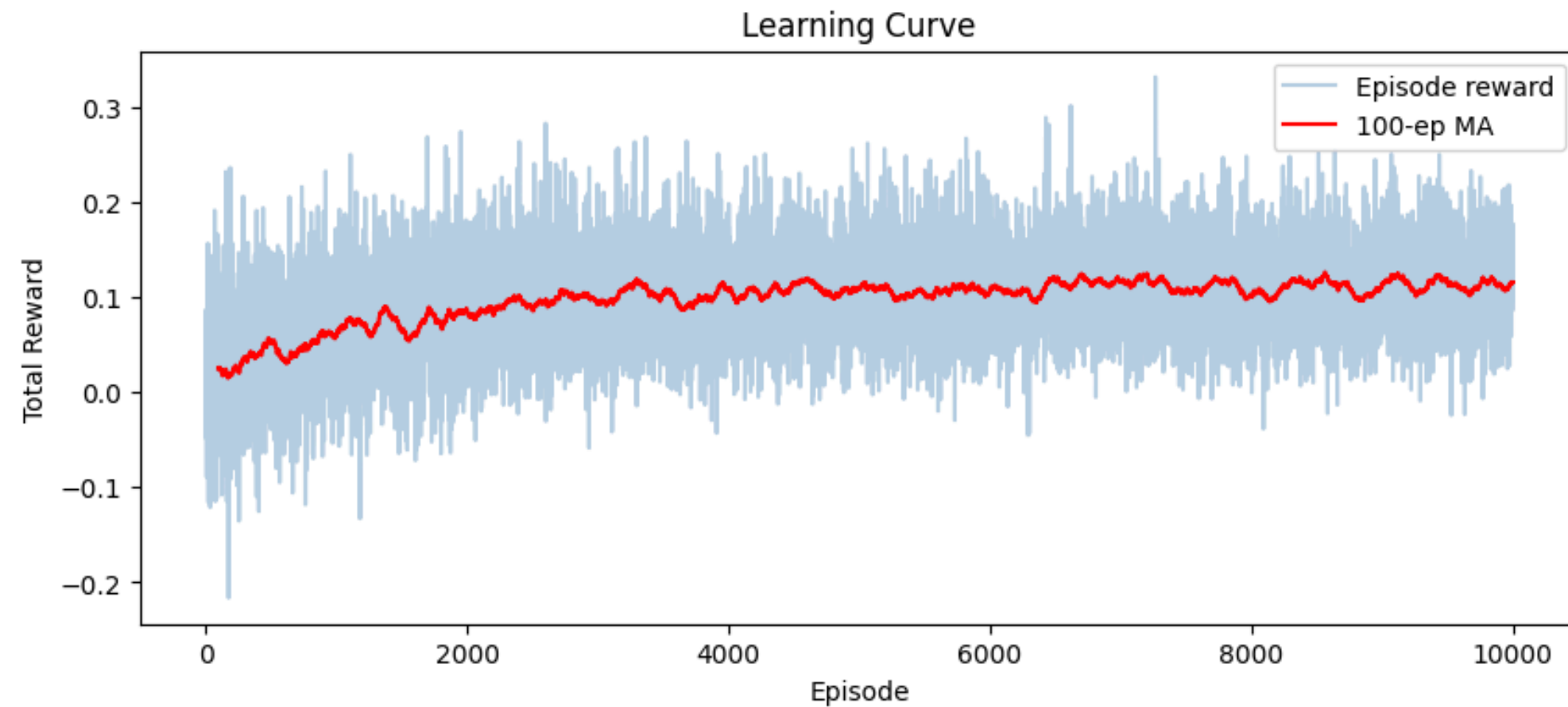
## 離散化

- Q学習では連続的な状態を扱えないため、価格は0.95 ~ 1.15の間を40分割している。

# 強化学習の手順

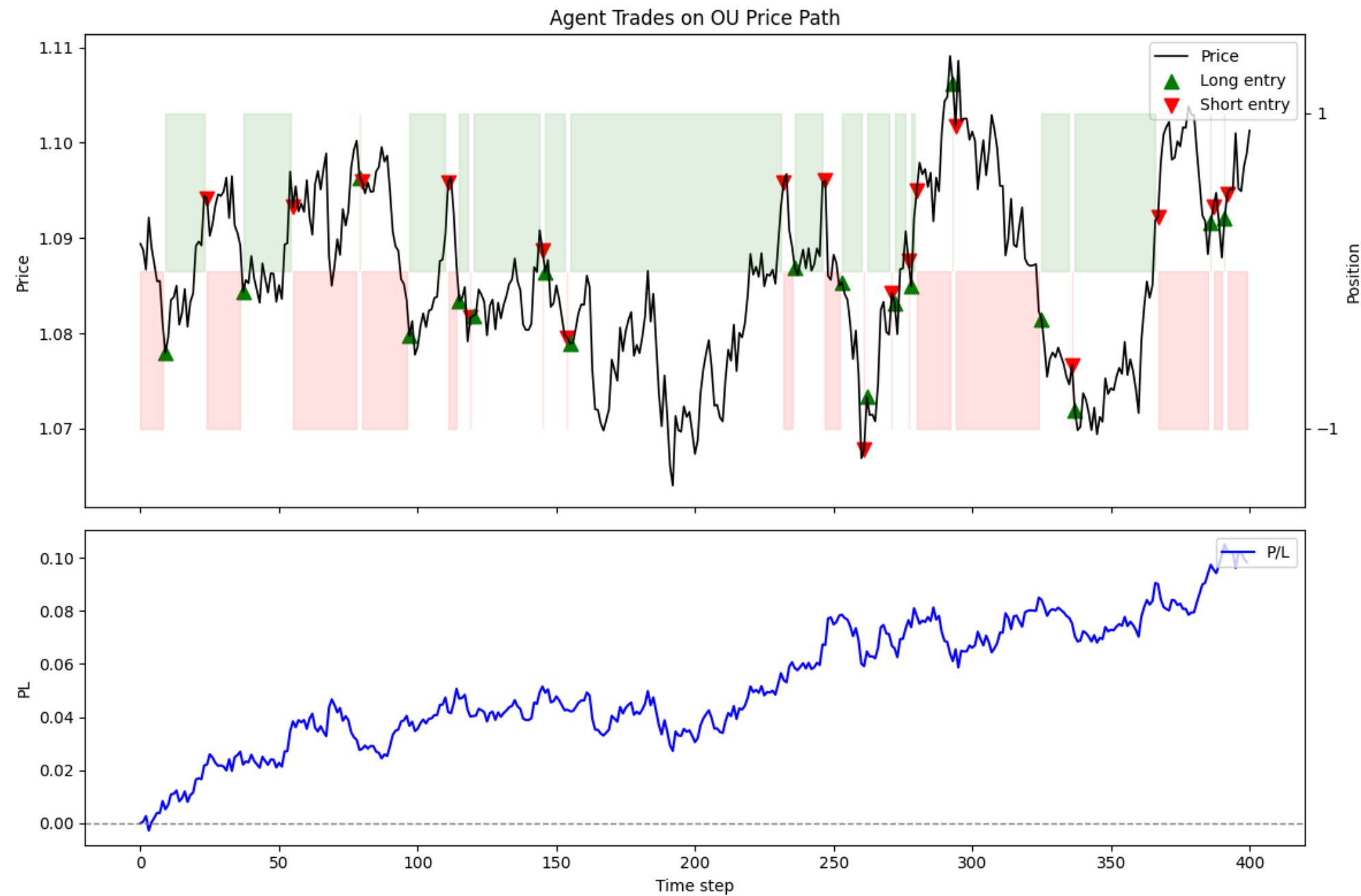
- 以下のパラメータに基づいて、10,000エピソード×500営業日学習した。

```
eps_start    = 0.3          # 初期  $\epsilon$   
eps_final    = 0.001        # 最終  $\epsilon$   
eps_decay    = 0.999        # 各エピソード後に掛ける係数  
alpha        = 0.05         # 学習率  
gamma        = 0.99         # 割引率
```



# テスト

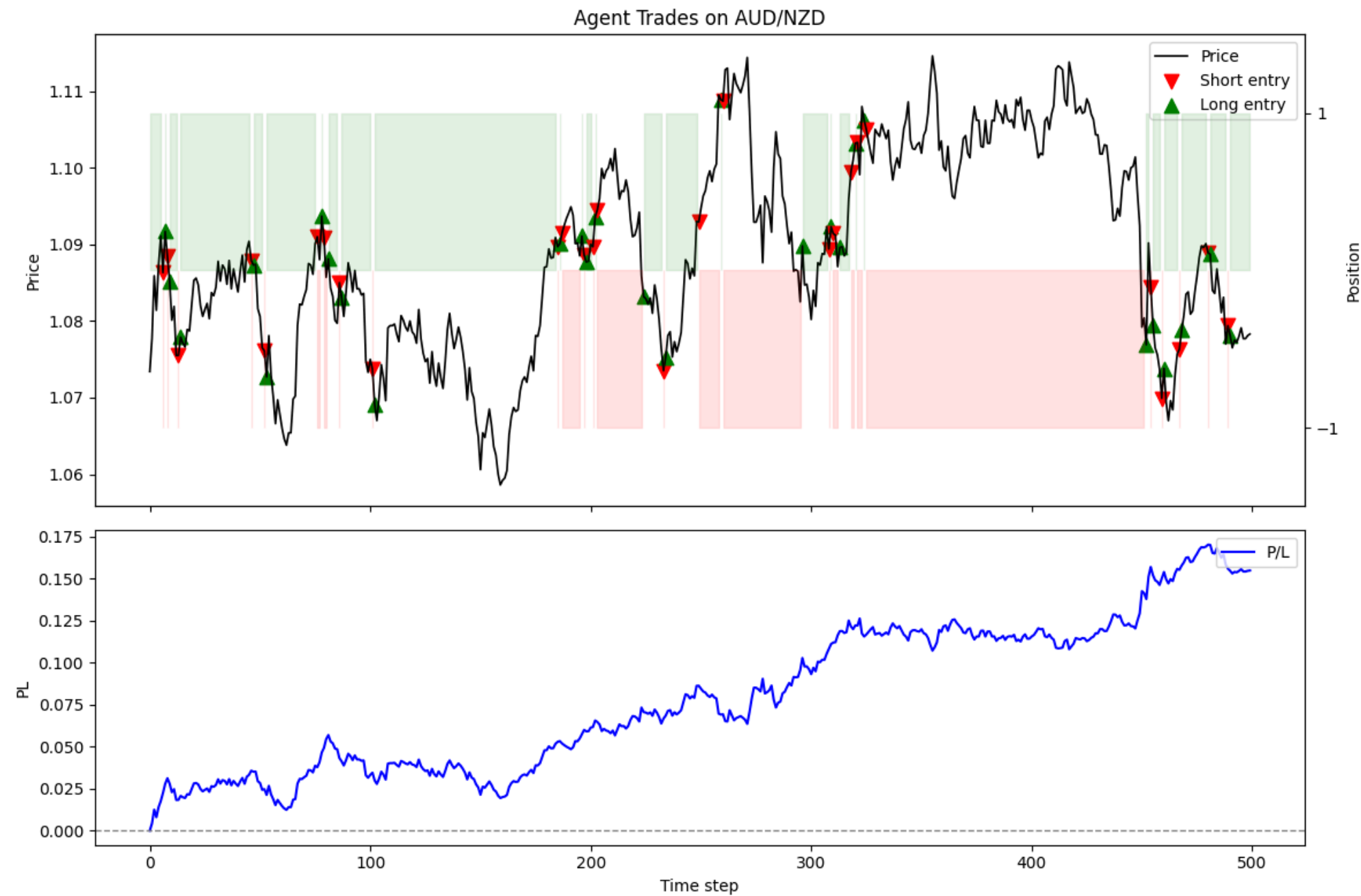
- PLは500営業日間1AUDをロングショートしたときの時価評価(仮想的な相場でのシミュレーション)
- 適当なOU過程に対し、テストするとおおよそ良好なパフォーマンスを出す様子が観察できる。





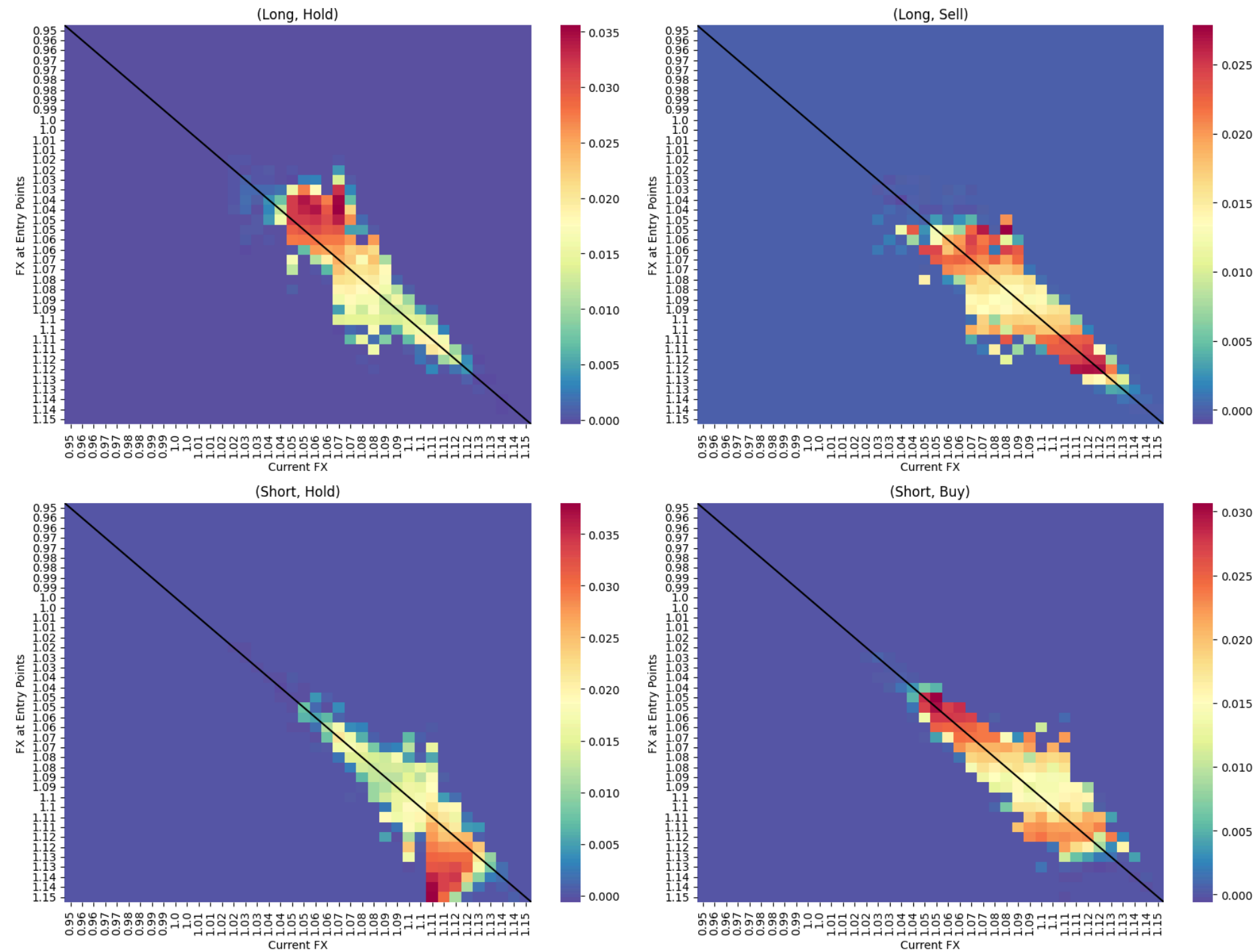
# 実データでバックテスト

- もし平均回帰性とその平均水準が確信できる場合はリターンが出る(期間は2023/7/14~2025/6/12)。
- 2024/10~2025/3など、異なる水準で平均回帰している期間では回帰を取りに行けない。



# Q-テーブルの可視化

- Q学習ではQ-テーブルと呼ばれる状態空間×行動空間の次元を持つ行列を更新することで学習し、これを用いて意思決定を行う。以下のヒートマップでは赤いものほど選択されやすい。



# Q-テーブルの観察

## ヒートマップの見方

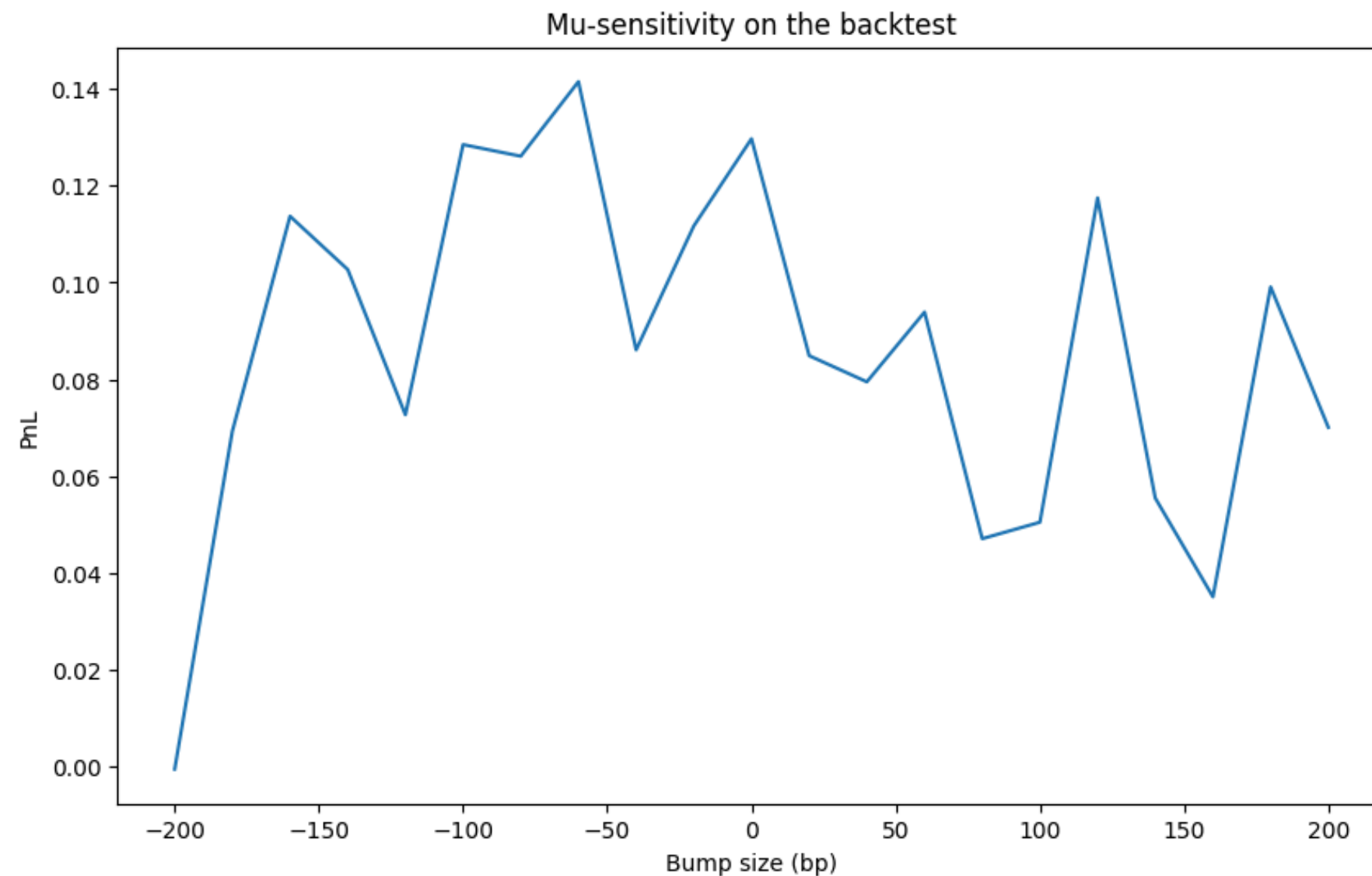
- グラフタイトルは (現在のポジション, そこで取る投資行動) を意味する。
- 縦軸はエントリー時の為替スポット、横軸は現在の為替スポットを表す。
- ロングのグラフでは右上が含み益、左下が含み損。ショートの場合は逆。

## 観察

- 低い価格でロングした場合は含み益が出てもしばらくポジションを閉じない(Holdの大きな赤)。ショートの場合、その逆。
- 利確は平均付近でちょっとと、深く逆側に振れてからのところで赤が濃い。

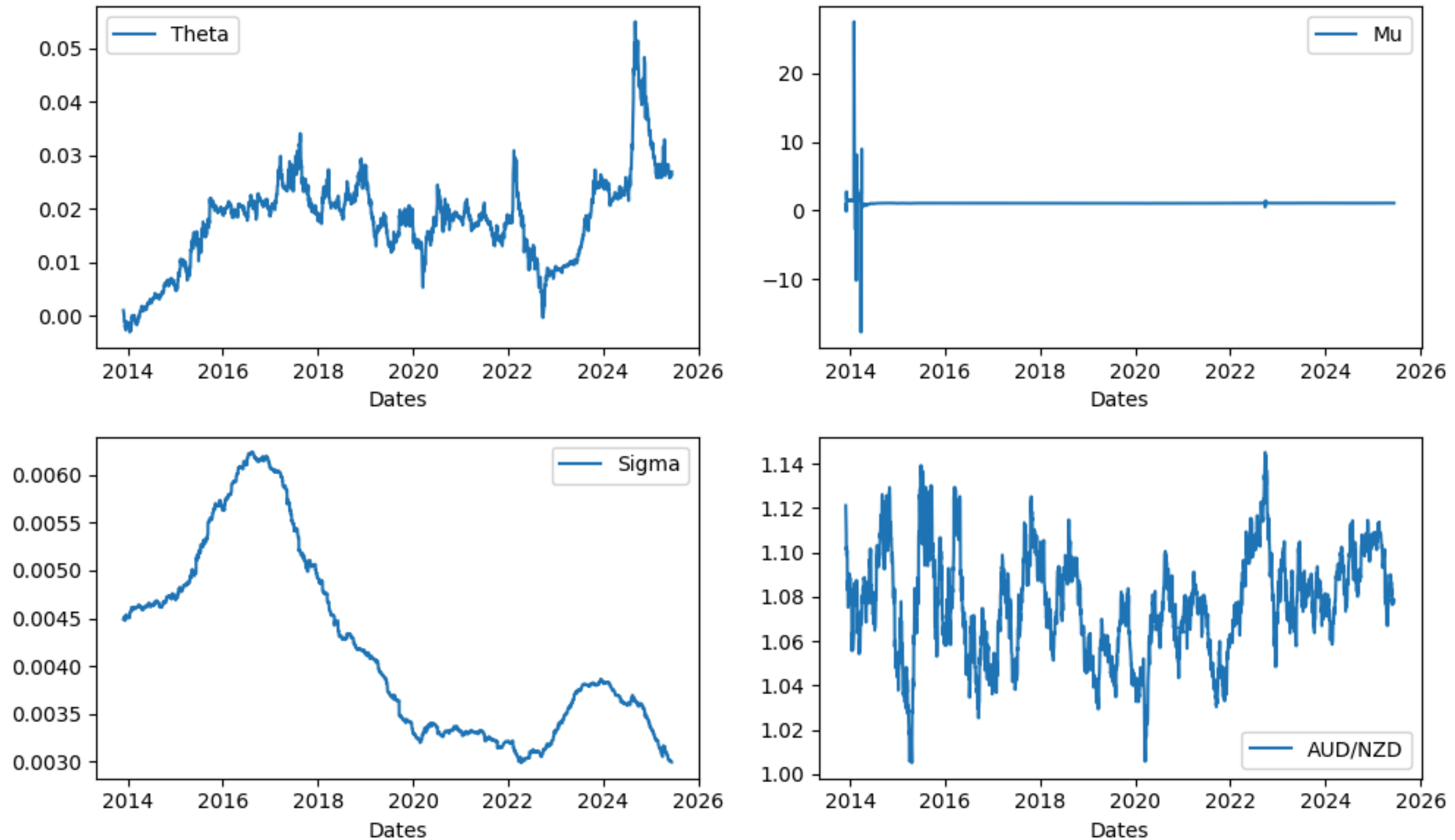
# 仮定する $\mu$ に対する感応度

- 平均を-200 bp ~ +200 bpに変動させたときのプロット。(学習のたびにそれぞれ出来が違うため本当は統計的な回数検証すべきだが、)大まかには平均の推定がずれるとパフォーマンスが低下することが見て取れる。
- 平均回帰は平均にベットするため、これをある程度正確に把握できていないとトレードで勝てない。



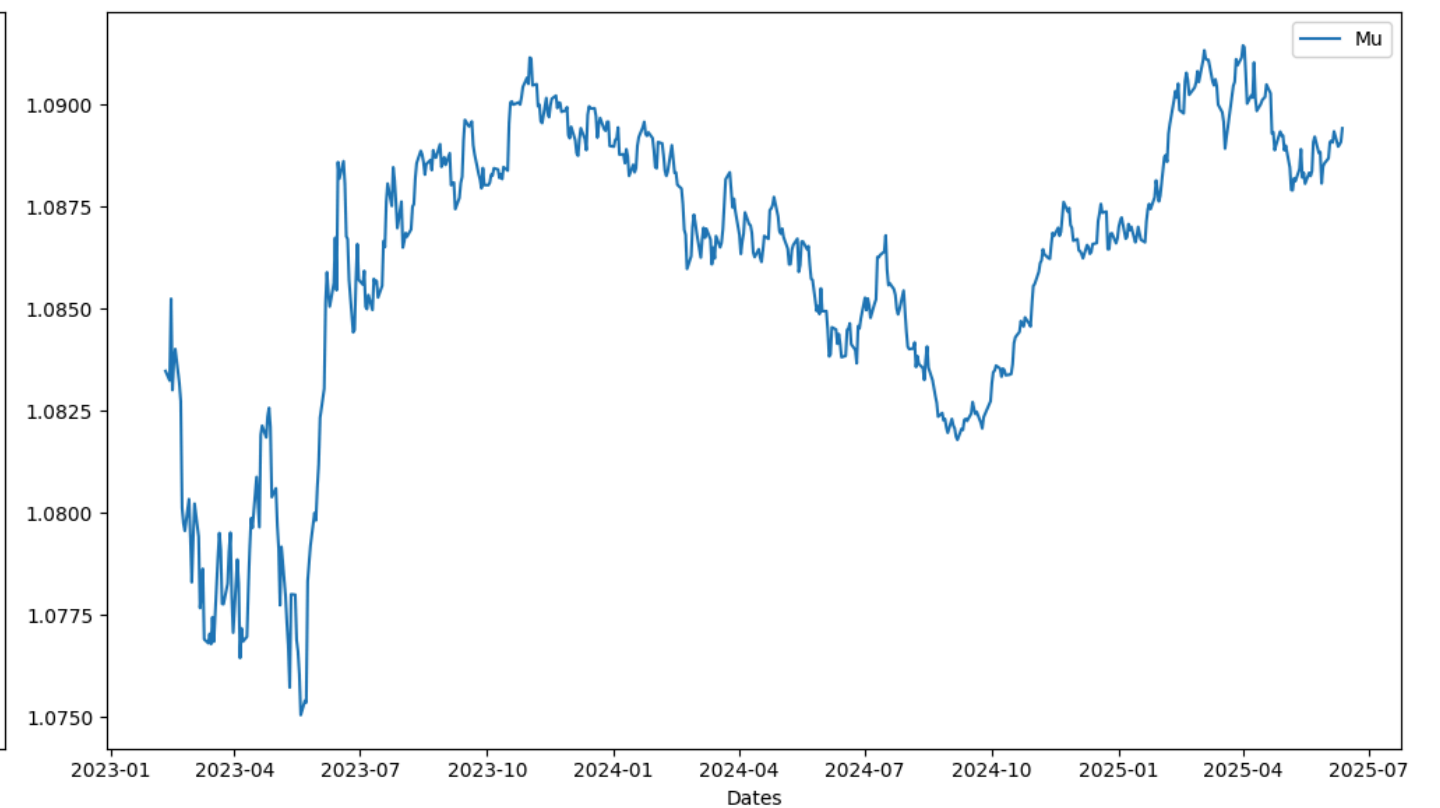
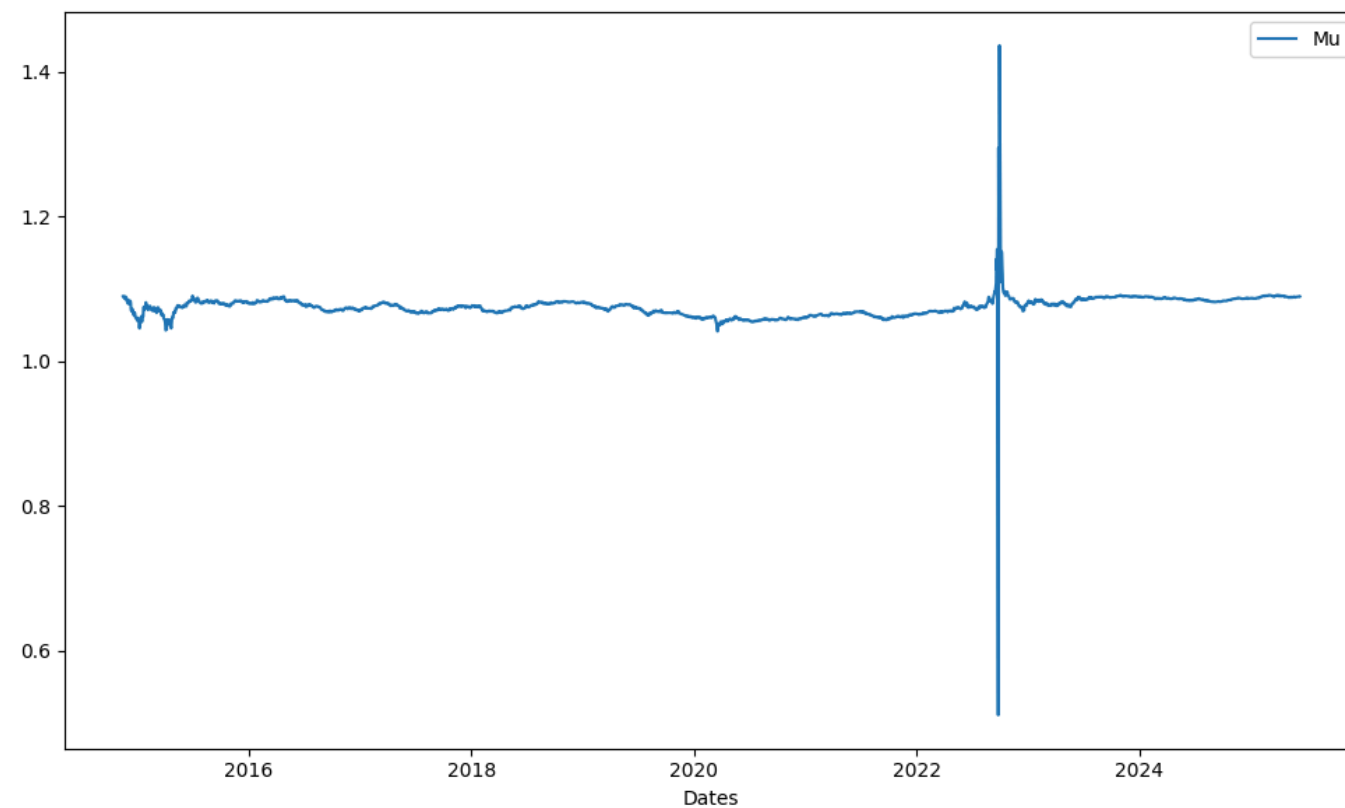
# 推定されるパラメータの定常性

- 過去10年間のについて500営業日ウィンドウでOUパラメータをキャリブレーションしたときの各パラメータの値をプロットした



# $\mu$ のグラフを拡大

- コロナショックを除き、かなり長い期間一定であるように見える。
- 安定した2023年以降は $\mu$ が150bpくらいの幅で推移しており、最大幅で外すとワークしなくなる可能性がある(例えば、2023/6~2023/12など)。



# 今後の方向性

---

- 上記のように「特定の相場状況のときに機能するトレードボット」を簡単にたくさん作れるので、もし需要があればいろんな相場に応じたボットを作ってバックテストしたり、トレーダーツールにしたい
- また今回はスポットしか見てないが、「特定のテクニカル/ファンダメだけを見るトレードボット」に拡張することも可能。現在USDJPYの過去データに対して、MA線やRSIを見てトレードするボットを実装中。
- 機械学習モデルの判断を人間が解釈して傾向分析をしたり、ルール化することでQISにも活用したい