

# train

August 21, 2023

```
[1]: import pandas as pd
from stable_baselines3.common.logger import configure

from finrl.agents.stablebaselines3.models import DRLAgent
from finrl.config import INDICATORS, TRAINED_MODEL_DIR, RESULTS_DIR
from finrl.main import check_and_make_directories
from finrl.meta.env_stock_trading.env_stocktrading import StockTradingEnv

check_and_make_directories([TRAINED_MODEL_DIR])
```

## 0.1 Read data

```
[2]: train = pd.read_csv('train_data.csv')
train = train.set_index(train.columns[0])
train.index.names = ['']
```

## 0.2 Construct the environment

```
[3]: stock_dimension = len(train.tic.unique())
state_space = 1 + 2*stock_dimension + len(INDICATORS)*stock_dimension
print(f"Stock Dimension: {stock_dimension}, State Space: {state_space}")
```

Stock Dimension: 29, State Space: 291

```
[4]: buy_cost_list = sell_cost_list = [0.001] * stock_dimension
num_stock_shares = [0] * stock_dimension

env_kwargs = {
    "hmax": 100,
    "initial_amount": 1000000,
    "num_stock_shares": num_stock_shares,
    "buy_cost_pct": buy_cost_list,
    "sell_cost_pct": sell_cost_list,
    "state_space": state_space,
    "stock_dim": stock_dimension,
    "tech_indicator_list": INDICATORS,
    "action_space": stock_dimension,
```

```

    "reward_scaling": 1e-4
}

e_train_gym = StockTradingEnv(df = train, **env_kwargs)

```

### 0.3 Environment for training

```

[5]: env_train, _ = e_train_gym.get_sb_env()
      print(type(env_train))

<class 'stable_baselines3.common.vec_env.dummy_vec_env.DummyVecEnv'>

```

## 1 Part 3: Train DRL Agents

```

[6]: agent = DRLAgent(env = env_train)

# Set the corresponding values to 'True' for the algorithms that you want to use
if_using_a2c = True
if_using_ddpg = True
if_using_ppo = True
if_using_td3 = True
if_using_sac = True

```

### 1.0.1 Agent 1: A2C

```

[7]: agent = DRLAgent(env = env_train)
      model_a2c = agent.get_model("a2c")

      if if_using_a2c:
          # set up logger
          tmp_path = RESULTS_DIR + '/a2c'
          new_logger_a2c = configure(tmp_path, ["stdout", "csv", "tensorboard"])
          # Set new logger
          model_a2c.set_logger(new_logger_a2c)

```

```

{'n_steps': 5, 'ent_coef': 0.01, 'learning_rate': 0.0007}
Using cuda device
Logging to results/a2c

```

```

[8]: trained_a2c = agent.train_model(model=model_a2c,
                                     tb_log_name='a2c',
                                     total_timesteps=50000) if if_using_a2c else None

```

```

-----
| time/                |                |

```

	fps	64	
	iterations	100	
	time_elapsed	7	
	total_timesteps	500	
	train/		
	entropy_loss	-41.2	
	explained_variance	0.282	
	learning_rate	0.0007	
	n_updates	99	
	policy_loss	38.4	
	reward	0.11025473	
	std	1	
	value_loss	1.09	

	time/		
	fps	76	
	iterations	200	
	time_elapsed	13	
	total_timesteps	1000	
	train/		
	entropy_loss	-41.2	
	explained_variance	-1.19e-07	
	learning_rate	0.0007	
	n_updates	199	
	policy_loss	-55.4	
	reward	-1.3854662	
	std	1	
	value_loss	2.56	

	time/		
	fps	81	
	iterations	300	
	time_elapsed	18	
	total_timesteps	1500	
	train/		
	entropy_loss	-41.3	
	explained_variance	0.0042	
	learning_rate	0.0007	
	n_updates	299	
	policy_loss	-339	
	reward	3.391866	
	std	1	
	value_loss	70.1	

	time/		
--	-------	--	--

	fps		86	
	iterations		400	
	time_elapsed		23	
	total_timesteps		2000	
	train/			
	entropy_loss		-41.3	
	explained_variance		5.96e-08	
	learning_rate		0.0007	
	n_updates		399	
	policy_loss		-70.7	
	reward		6.332755	
	std		1	
	value_loss		4.11	

	time/			
	fps		87	
	iterations		500	
	time_elapsed		28	
	total_timesteps		2500	
	train/			
	entropy_loss		-41.3	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		499	
	policy_loss		380	
	reward		-4.429727	
	std		1.01	
	value_loss		104	

	time/			
	fps		88	
	iterations		600	
	time_elapsed		33	
	total_timesteps		3000	
	train/			
	entropy_loss		-41.3	
	explained_variance		-0.00171	
	learning_rate		0.0007	
	n_updates		599	
	policy_loss		193	
	reward		0.087448716	
	std		1.01	
	value_loss		22.7	

	time/			
--	-------	--	--	--

	fps	89	
	iterations	700	
	time_elapsed	39	
	total_timesteps	3500	
	train/		
	entropy_loss	-41.3	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	699	
	policy_loss	-65.8	
	reward	-2.7318516	
	std	1	
	value_loss	2.78	

	time/		
	fps	89	
	iterations	800	
	time_elapsed	44	
	total_timesteps	4000	
	train/		
	entropy_loss	-41.3	
	explained_variance	1.19e-07	
	learning_rate	0.0007	
	n_updates	799	
	policy_loss	-122	
	reward	-3.5182903	
	std	1.01	
	value_loss	12.1	

	time/		
	fps	89	
	iterations	900	
	time_elapsed	50	
	total_timesteps	4500	
	train/		
	entropy_loss	-41.4	
	explained_variance	-1.19e-07	
	learning_rate	0.0007	
	n_updates	899	
	policy_loss	81.2	
	reward	-0.7624608	
	std	1.01	
	value_loss	4.53	

	time/		
--	-------	--	--

	fps	90	
	iterations	1000	
	time_elapsed	55	
	total_timesteps	5000	
	train/		
	entropy_loss	-41.4	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	999	
	policy_loss	90.2	
	reward	-0.76001424	
	std	1.01	
	value_loss	5.31	

	time/		
	fps	91	
	iterations	1100	
	time_elapsed	60	
	total_timesteps	5500	
	train/		
	entropy_loss	-41.4	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	1099	
	policy_loss	-98.2	
	reward	0.8892505	
	std	1.01	
	value_loss	8.8	

	time/		
	fps	92	
	iterations	1200	
	time_elapsed	64	
	total_timesteps	6000	
	train/		
	entropy_loss	-41.4	
	explained_variance	-0.0536	
	learning_rate	0.0007	
	n_updates	1199	
	policy_loss	-72.2	
	reward	0.26658863	
	std	1.01	
	value_loss	3.88	

	time/		
--	-------	--	--

	fps	93	
	iterations	1300	
	time_elapsed	69	
	total_timesteps	6500	
	train/		
	entropy_loss	-41.5	
	explained_variance	0.0217	
	learning_rate	0.0007	
	n_updates	1299	
	policy_loss	-1.81	
	reward	-2.2109706	
	std	1.01	
	value_loss	1.03	

---

	time/		
	fps	94	
	iterations	1400	
	time_elapsed	74	
	total_timesteps	7000	
	train/		
	entropy_loss	-41.5	
	explained_variance	5.96e-08	
	learning_rate	0.0007	
	n_updates	1399	
	policy_loss	130	
	reward	0.8835194	
	std	1.01	
	value_loss	12.6	

---

	time/		
	fps	94	
	iterations	1500	
	time_elapsed	79	
	total_timesteps	7500	
	train/		
	entropy_loss	-41.5	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	1499	
	policy_loss	-186	
	reward	1.2555037	
	std	1.01	
	value_loss	45.1	

---

	time/		
--	-------	--	--

	fps	95	
	iterations	1600	
	time_elapsed	83	
	total_timesteps	8000	
	train/		
	entropy_loss	-41.5	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	1599	
	policy_loss	251	
	reward	0.24808238	
	std	1.01	
	value_loss	37	

	time/		
	fps	96	
	iterations	1700	
	time_elapsed	88	
	total_timesteps	8500	
	train/		
	entropy_loss	-41.5	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	1699	
	policy_loss	-19.5	
	reward	7.7073183	
	std	1.01	
	value_loss	7.66	

	time/		
	fps	96	
	iterations	1800	
	time_elapsed	93	
	total_timesteps	9000	
	train/		
	entropy_loss	-41.5	
	explained_variance	0.000519	
	learning_rate	0.0007	
	n_updates	1799	
	policy_loss	-30.7	
	reward	0.51203763	
	std	1.01	
	value_loss	0.869	

	time/		
--	-------	--	--



	fps		97	
	iterations		1900	
	time_elapsed		97	
	total_timesteps		9500	
	train/			
	entropy_loss		-41.5	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		1899	
	policy_loss		44	
	reward		0.041149914	
	std		1.01	
	value_loss		1.63	

---

	time/			
	fps		97	
	iterations		2000	
	time_elapsed		102	
	total_timesteps		10000	
	train/			
	entropy_loss		-41.6	
	explained_variance		1.19e-07	
	learning_rate		0.0007	
	n_updates		1999	
	policy_loss		10.6	
	reward		0.7703739	
	std		1.02	
	value_loss		0.256	

---

	time/			
	fps		98	
	iterations		2100	
	time_elapsed		106	
	total_timesteps		10500	
	train/			
	entropy_loss		-41.6	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		2099	
	policy_loss		32.8	
	reward		0.6852897	
	std		1.02	
	value_loss		2.1	

---

	time/			
--	-------	--	--	--

	fps	98	
	iterations	2200	
	time_elapsed	111	
	total_timesteps	11000	
	train/		
	entropy_loss	-41.6	
	explained_variance	-1.19e-07	
	learning_rate	0.0007	
	n_updates	2199	
	policy_loss	-32.5	
	reward	-7.1390004	
	std	1.02	
	value_loss	6.56	

	time/		
	fps	99	
	iterations	2300	
	time_elapsed	115	
	total_timesteps	11500	
	train/		
	entropy_loss	-41.6	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	2299	
	policy_loss	-2.37e+03	
	reward	-10.550006	
	std	1.02	
	value_loss	3.2e+03	

	time/		
	fps	99	
	iterations	2400	
	time_elapsed	120	
	total_timesteps	12000	
	train/		
	entropy_loss	-41.6	
	explained_variance	0.0284	
	learning_rate	0.0007	
	n_updates	2399	
	policy_loss	67.5	
	reward	0.3746248	
	std	1.02	
	value_loss	6.68	

	time/		
--	-------	--	--

	fps	99	
	iterations	2500	
	time_elapsed	125	
	total_timesteps	12500	
	train/		
	entropy_loss	-41.7	
	explained_variance	-1.19e-07	
	learning_rate	0.0007	
	n_updates	2499	
	policy_loss	-30.5	
	reward	-0.6775755	
	std	1.02	
	value_loss	1.23	

---

	time/		
	fps	100	
	iterations	2600	
	time_elapsed	129	
	total_timesteps	13000	
	train/		
	entropy_loss	-41.7	
	explained_variance	-1.19e-07	
	learning_rate	0.0007	
	n_updates	2599	
	policy_loss	32.3	
	reward	3.0196373	
	std	1.02	
	value_loss	1.2	

---

	time/		
	fps	100	
	iterations	2700	
	time_elapsed	134	
	total_timesteps	13500	
	train/		
	entropy_loss	-41.7	
	explained_variance	0.0648	
	learning_rate	0.0007	
	n_updates	2699	
	policy_loss	48.5	
	reward	-1.8973987	
	std	1.02	
	value_loss	1.88	

---

	time/		
--	-------	--	--

	fps		100	
	iterations		2800	
	time_elapsed		139	
	total_timesteps		14000	
	train/			
	entropy_loss		-41.7	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		2799	
	policy_loss		308	
	reward		2.8944852	
	std		1.02	
	value_loss		57.3	

	time/			
	fps		101	
	iterations		2900	
	time_elapsed		143	
	total_timesteps		14500	
	train/			
	entropy_loss		-41.7	
	explained_variance		-3.58e-07	
	learning_rate		0.0007	
	n_updates		2899	
	policy_loss		-117	
	reward		1.8488793	
	std		1.02	
	value_loss		8.88	

	time/			
	fps		101	
	iterations		3000	
	time_elapsed		148	
	total_timesteps		15000	
	train/			
	entropy_loss		-41.8	
	explained_variance		5.96e-08	
	learning_rate		0.0007	
	n_updates		2999	
	policy_loss		3.98	
	reward		0.260266	
	std		1.02	
	value_loss		0.102	

	time/			
--	-------	--	--	--

	fps		101	
	iterations		3100	
	time_elapsed		152	
	total_timesteps		15500	
	train/			
	entropy_loss		-41.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		3099	
	policy_loss		6.14	
	reward		-0.18214704	
	std		1.02	
	value_loss		0.58	

	time/			
	fps		101	
	iterations		3200	
	time_elapsed		157	
	total_timesteps		16000	
	train/			
	entropy_loss		-41.9	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		3199	
	policy_loss		-98.9	
	reward		-1.9565057	
	std		1.02	
	value_loss		19.1	

	time/			
	fps		101	
	iterations		3300	
	time_elapsed		162	
	total_timesteps		16500	
	train/			
	entropy_loss		-41.9	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		3299	
	policy_loss		-199	
	reward		1.7313249	
	std		1.03	
	value_loss		32.1	

	time/			
--	-------	--	--	--

	fps		101	
	iterations		3400	
	time_elapsed		166	
	total_timesteps		17000	
	train/			
	entropy_loss		-41.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		3399	
	policy_loss		120	
	reward		5.4731264	
	std		1.02	
	value_loss		16.4	

	time/			
	fps		102	
	iterations		3500	
	time_elapsed		171	
	total_timesteps		17500	
	train/			
	entropy_loss		-41.8	
	explained_variance		-0.027	
	learning_rate		0.0007	
	n_updates		3499	
	policy_loss		111	
	reward		0.15176553	
	std		1.02	
	value_loss		8.87	

	time/			
	fps		102	
	iterations		3600	
	time_elapsed		176	
	total_timesteps		18000	
	train/			
	entropy_loss		-41.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		3599	
	policy_loss		-22.9	
	reward		1.2557095	
	std		1.02	
	value_loss		1.16	

	time/			
--	-------	--	--	--

	fps	102	
	iterations	3700	
	time_elapsed	180	
	total_timesteps	18500	
	train/		
	entropy_loss	-41.9	
	explained_variance	5.96e-08	
	learning_rate	0.0007	
	n_updates	3699	
	policy_loss	112	
	reward	-1.9729638	
	std	1.03	
	value_loss	8.67	

---

	time/		
	fps	102	
	iterations	3800	
	time_elapsed	185	
	total_timesteps	19000	
	train/		
	entropy_loss	-41.9	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	3799	
	policy_loss	120	
	reward	0.6203998	
	std	1.03	
	value_loss	8.83	

---

	time/		
	fps	102	
	iterations	3900	
	time_elapsed	190	
	total_timesteps	19500	
	train/		
	entropy_loss	-42.1	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	3899	
	policy_loss	-217	
	reward	1.0554643	
	std	1.03	
	value_loss	29.7	

---

	time/		
--	-------	--	--

	fps		102	
	iterations		4000	
	time_elapsed		194	
	total_timesteps		20000	
	train/			
	entropy_loss		-42.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		3999	
	policy_loss		-314	
	reward		4.7121725	
	std		1.03	
	value_loss		54.4	

	time/			
	fps		102	
	iterations		4100	
	time_elapsed		199	
	total_timesteps		20500	
	train/			
	entropy_loss		-42.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		4099	
	policy_loss		-7.29	
	reward		0.08846258	
	std		1.03	
	value_loss		0.762	

	time/			
	fps		102	
	iterations		4200	
	time_elapsed		204	
	total_timesteps		21000	
	train/			
	entropy_loss		-42.1	
	explained_variance		1.19e-07	
	learning_rate		0.0007	
	n_updates		4199	
	policy_loss		-85.1	
	reward		0.9444181	
	std		1.03	
	value_loss		5.96	

	time/			
--	-------	--	--	--



	fps	102	
	iterations	4300	
	time_elapsed	208	
	total_timesteps	21500	
	train/		
	entropy_loss	-42.1	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	4299	
	policy_loss	-134	
	reward	4.770794	
	std	1.03	
	value_loss	12.3	

	time/		
	fps	102	
	iterations	4400	
	time_elapsed	213	
	total_timesteps	22000	
	train/		
	entropy_loss	-42.1	
	explained_variance	-1.19e-07	
	learning_rate	0.0007	
	n_updates	4399	
	policy_loss	105	
	reward	1.9008546	
	std	1.03	
	value_loss	11.8	

	time/		
	fps	103	
	iterations	4500	
	time_elapsed	218	
	total_timesteps	22500	
	train/		
	entropy_loss	-42.1	
	explained_variance	1.19e-07	
	learning_rate	0.0007	
	n_updates	4499	
	policy_loss	206	
	reward	-1.3398763	
	std	1.03	
	value_loss	37.9	

	time/		
--	-------	--	--

	fps		103	
	iterations		4600	
	time_elapsed		222	
	total_timesteps		23000	
	train/			
	entropy_loss		-42.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		4599	
	policy_loss		168	
	reward		6.0184593	
	std		1.03	
	value_loss		22.4	

	time/			
	fps		103	
	iterations		4700	
	time_elapsed		227	
	total_timesteps		23500	
	train/			
	entropy_loss		-42.2	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		4699	
	policy_loss		-110	
	reward		0.3417877	
	std		1.04	
	value_loss		9.51	

	time/			
	fps		103	
	iterations		4800	
	time_elapsed		232	
	total_timesteps		24000	
	train/			
	entropy_loss		-42.3	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		4799	
	policy_loss		-108	
	reward		-0.976139	
	std		1.04	
	value_loss		6.47	

	time/			
--	-------	--	--	--

	fps	103	
	iterations	4900	
	time_elapsed	237	
	total_timesteps	24500	
	train/		
	entropy_loss	-42.3	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	4899	
	policy_loss	-36.5	
	reward	0.47295007	
	std	1.04	
	value_loss	1.31	

---

	time/		
	fps	103	
	iterations	5000	
	time_elapsed	241	
	total_timesteps	25000	
	train/		
	entropy_loss	-42.4	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	4999	
	policy_loss	56.9	
	reward	-0.5308896	
	std	1.04	
	value_loss	4.45	

---

	time/		
	fps	103	
	iterations	5100	
	time_elapsed	246	
	total_timesteps	25500	
	train/		
	entropy_loss	-42.4	
	explained_variance	0	
	learning_rate	0.0007	
	n_updates	5099	
	policy_loss	-110	
	reward	-1.9031373	
	std	1.04	
	value_loss	17.3	

---

	time/		
--	-------	--	--

	fps		103	
	iterations		5200	
	time_elapsed		251	
	total_timesteps		26000	
	train/			
	entropy_loss		-42.4	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		5199	
	policy_loss		-480	
	reward		3.9432917	
	std		1.04	
	value_loss		147	

-----

day: 2892, episode: 10  
begin\_total\_asset: 1000000.00  
end\_total\_asset: 3758073.56  
total\_reward: 2758073.56  
total\_cost: 5975.66  
total\_trades: 49683  
Sharpe: 0.809

=====

	time/			
	fps		103	
	iterations		5300	
	time_elapsed		256	
	total_timesteps		26500	
	train/			
	entropy_loss		-42.4	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		5299	
	policy_loss		-27.7	
	reward		0.5097055	
	std		1.04	
	value_loss		0.468	

-----

	time/			
	fps		103	
	iterations		5400	
	time_elapsed		260	
	total_timesteps		27000	
	train/			
	entropy_loss		-42.4	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		5399	
	policy_loss		97.4	
	reward		-0.21185152	
	std		1.04	
	value_loss		13.7	

---

	time/			
	fps		103	
	iterations		5500	
	time_elapsed		265	
	total_timesteps		27500	
	train/			
	entropy_loss		-42.4	
	explained_variance		0.00356	
	learning_rate		0.0007	
	n_updates		5499	
	policy_loss		-55.4	
	reward		2.0259786	
	std		1.04	
	value_loss		19.7	

---

	time/			
	fps		103	
	iterations		5600	
	time_elapsed		269	
	total_timesteps		28000	
	train/			
	entropy_loss		-42.4	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		5599	
	policy_loss		-113	
	reward		-1.3711272	
	std		1.04	
	value_loss		10.9	

---

	time/			
	fps		103	
	iterations		5700	
	time_elapsed		274	
	total_timesteps		28500	
	train/			
	entropy_loss		-42.4	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		5699	
	policy_loss		-72	
	reward		-0.98385024	
	std		1.04	
	value_loss		7.47	

	time/			
	fps		103	
	iterations		5800	
	time_elapsed		279	
	total_timesteps		29000	
	train/			
	entropy_loss		-42.4	
	explained_variance		-0.126	
	learning_rate		0.0007	
	n_updates		5799	
	policy_loss		-34.8	
	reward		1.0769047	
	std		1.04	
	value_loss		2.4	

	time/			
	fps		103	
	iterations		5900	
	time_elapsed		283	
	total_timesteps		29500	
	train/			
	entropy_loss		-42.3	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		5899	
	policy_loss		-16.5	
	reward		-0.034104265	
	std		1.04	
	value_loss		0.287	

	time/			
	fps		103	
	iterations		6000	
	time_elapsed		288	
	total_timesteps		30000	
	train/			
	entropy_loss		-42.4	
	explained_variance		5.96e-08	
	learning_rate		0.0007	

	n_updates		5999	
	policy_loss		58	
	reward		-0.38149226	
	std		1.04	
	value_loss		3.77	

	time/			
	fps		104	
	iterations		6100	
	time_elapsed		293	
	total_timesteps		30500	
	train/			
	entropy_loss		-42.4	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		6099	
	policy_loss		35.5	
	reward		-1.8691139	
	std		1.05	
	value_loss		4.74	

	time/			
	fps		104	
	iterations		6200	
	time_elapsed		297	
	total_timesteps		31000	
	train/			
	entropy_loss		-42.5	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		6199	
	policy_loss		-9	
	reward		-0.48611382	
	std		1.05	
	value_loss		2.04	

	time/			
	fps		104	
	iterations		6300	
	time_elapsed		302	
	total_timesteps		31500	
	train/			
	entropy_loss		-42.5	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		6299	
	policy_loss		41	
	reward		3.1559389	
	std		1.05	
	value_loss		5.11	

	time/			
	fps		104	
	iterations		6400	
	time_elapsed		307	
	total_timesteps		32000	
	train/			
	entropy_loss		-42.6	
	explained_variance		0.0421	
	learning_rate		0.0007	
	n_updates		6399	
	policy_loss		49.2	
	reward		0.8848122	
	std		1.05	
	value_loss		1.73	

	time/			
	fps		104	
	iterations		6500	
	time_elapsed		311	
	total_timesteps		32500	
	train/			
	entropy_loss		-42.6	
	explained_variance		0.289	
	learning_rate		0.0007	
	n_updates		6499	
	policy_loss		0.413	
	reward		-3.396548	
	std		1.05	
	value_loss		1	

	time/			
	fps		104	
	iterations		6600	
	time_elapsed		316	
	total_timesteps		33000	
	train/			
	entropy_loss		-42.7	
	explained_variance		0	
	learning_rate		0.0007	



	n_updates		6599	
	policy_loss		-79.3	
	reward		-0.46146256	
	std		1.05	
	value_loss		6.69	

---

	time/			
	fps		104	
	iterations		6700	
	time_elapsed		321	
	total_timesteps		33500	
	train/			
	entropy_loss		-42.7	
	explained_variance		0.00458	
	learning_rate		0.0007	
	n_updates		6699	
	policy_loss		-780	
	reward		-5.730236	
	std		1.06	
	value_loss		382	

---

	time/			
	fps		104	
	iterations		6800	
	time_elapsed		325	
	total_timesteps		34000	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		6799	
	policy_loss		-209	
	reward		0.55269736	
	std		1.06	
	value_loss		27.4	

---

	time/			
	fps		104	
	iterations		6900	
	time_elapsed		330	
	total_timesteps		34500	
	train/			
	entropy_loss		-42.7	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		6899	
	policy_loss		125	
	reward		-0.7730595	
	std		1.06	
	value_loss		24.3	

	time/			
	fps		104	
	iterations		7000	
	time_elapsed		335	
	total_timesteps		35000	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		6999	
	policy_loss		29.8	
	reward		0.17269997	
	std		1.06	
	value_loss		0.839	

	time/			
	fps		104	
	iterations		7100	
	time_elapsed		339	
	total_timesteps		35500	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		7099	
	policy_loss		24.9	
	reward		0.5368473	
	std		1.06	
	value_loss		0.69	

	time/			
	fps		104	
	iterations		7200	
	time_elapsed		344	
	total_timesteps		36000	
	train/			
	entropy_loss		-42.8	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	

	n_updates		7199	
	policy_loss		-283	
	reward		0.23633873	
	std		1.06	
	value_loss		40	

	time/			
	fps		104	
	iterations		7300	
	time_elapsed		349	
	total_timesteps		36500	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		7299	
	policy_loss		175	
	reward		-0.2561297	
	std		1.06	
	value_loss		25.8	

	time/			
	fps		104	
	iterations		7400	
	time_elapsed		353	
	total_timesteps		37000	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		7399	
	policy_loss		-18	
	reward		-5.361117	
	std		1.06	
	value_loss		6.32	

	time/			
	fps		104	
	iterations		7500	
	time_elapsed		358	
	total_timesteps		37500	
	train/			
	entropy_loss		-42.8	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	

	n_updates		7499	
	policy_loss		95.7	
	reward		1.2376721	
	std		1.06	
	value_loss		6.68	

	time/			
	fps		104	
	iterations		7600	
	time_elapsed		363	
	total_timesteps		38000	
	train/			
	entropy_loss		-42.7	
	explained_variance		5.96e-08	
	learning_rate		0.0007	
	n_updates		7599	
	policy_loss		-102	
	reward		1.2518629	
	std		1.06	
	value_loss		5.92	

	time/			
	fps		104	
	iterations		7700	
	time_elapsed		369	
	total_timesteps		38500	
	train/			
	entropy_loss		-42.8	
	explained_variance		1.19e-07	
	learning_rate		0.0007	
	n_updates		7699	
	policy_loss		-53.2	
	reward		1.7100748	
	std		1.06	
	value_loss		2.17	

	time/			
	fps		103	
	iterations		7800	
	time_elapsed		375	
	total_timesteps		39000	
	train/			
	entropy_loss		-42.8	
	explained_variance		1.19e-07	
	learning_rate		0.0007	

	n_updates		7799	
	policy_loss		-35.6	
	reward		-3.2164109	
	std		1.06	
	value_loss		1.91	

---

	time/			
	fps		103	
	iterations		7900	
	time_elapsed		380	
	total_timesteps		39500	
	train/			
	entropy_loss		-42.8	
	explained_variance		5.96e-08	
	learning_rate		0.0007	
	n_updates		7899	
	policy_loss		374	
	reward		1.6105194	
	std		1.06	
	value_loss		75.7	

---

	time/			
	fps		103	
	iterations		8000	
	time_elapsed		385	
	total_timesteps		40000	
	train/			
	entropy_loss		-42.8	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		7999	
	policy_loss		-19.7	
	reward		2.5263324	
	std		1.06	
	value_loss		1.82	

---

	time/			
	fps		103	
	iterations		8100	
	time_elapsed		390	
	total_timesteps		40500	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		8099	
	policy_loss		-40.7	
	reward		4.778373	
	std		1.06	
	value_loss		11.6	

	time/			
	fps		103	
	iterations		8200	
	time_elapsed		395	
	total_timesteps		41000	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		8199	
	policy_loss		39.1	
	reward		0.24881849	
	std		1.06	
	value_loss		1.11	

	time/			
	fps		103	
	iterations		8300	
	time_elapsed		399	
	total_timesteps		41500	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		8299	
	policy_loss		69.8	
	reward		-1.1919031	
	std		1.06	
	value_loss		3.5	

	time/			
	fps		103	
	iterations		8400	
	time_elapsed		404	
	total_timesteps		42000	
	train/			
	entropy_loss		-42.8	
	explained_variance		5.96e-08	
	learning_rate		0.0007	

	n_updates		8399	
	policy_loss		-6.27	
	reward		-1.4026177	
	std		1.06	
	value_loss		0.395	

	time/			
	fps		103	
	iterations		8500	
	time_elapsed		409	
	total_timesteps		42500	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		8499	
	policy_loss		-213	
	reward		-0.24122107	
	std		1.06	
	value_loss		26.6	

	time/			
	fps		103	
	iterations		8600	
	time_elapsed		413	
	total_timesteps		43000	
	train/			
	entropy_loss		-42.8	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		8599	
	policy_loss		48.2	
	reward		-3.835304	
	std		1.06	
	value_loss		38.2	

	time/			
	fps		103	
	iterations		8700	
	time_elapsed		418	
	total_timesteps		43500	
	train/			
	entropy_loss		-42.9	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		8699	
	policy_loss		2.8	
	reward		0.42422602	
	std		1.06	
	value_loss		0.954	

	time/			
	fps		103	
	iterations		8800	
	time_elapsed		423	
	total_timesteps		44000	
	train/			
	entropy_loss		-42.9	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		8799	
	policy_loss		-77.4	
	reward		1.1890944	
	std		1.07	
	value_loss		3.23	

	time/			
	fps		104	
	iterations		8900	
	time_elapsed		427	
	total_timesteps		44500	
	train/			
	entropy_loss		-43	
	explained_variance		5.96e-08	
	learning_rate		0.0007	
	n_updates		8899	
	policy_loss		79	
	reward		-0.68389106	
	std		1.07	
	value_loss		4.92	

	time/			
	fps		104	
	iterations		9000	
	time_elapsed		432	
	total_timesteps		45000	
	train/			
	entropy_loss		-43	
	explained_variance		0	
	learning_rate		0.0007	



	n_updates		8999	
	policy_loss		16.2	
	reward		-0.44394338	
	std		1.07	
	value_loss		4.04	

---

	time/			
	fps		104	
	iterations		9100	
	time_elapsed		437	
	total_timesteps		45500	
	train/			
	entropy_loss		-43.1	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	
	n_updates		9099	
	policy_loss		-21.4	
	reward		6.438101	
	std		1.07	
	value_loss		0.743	

---

	time/			
	fps		104	
	iterations		9200	
	time_elapsed		441	
	total_timesteps		46000	
	train/			
	entropy_loss		-43.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		9199	
	policy_loss		121	
	reward		1.0236892	
	std		1.07	
	value_loss		10.9	

---

	time/			
	fps		104	
	iterations		9300	
	time_elapsed		446	
	total_timesteps		46500	
	train/			
	entropy_loss		-43.1	
	explained_variance		-0.29	
	learning_rate		0.0007	

	n_updates		9299	
	policy_loss		-69	
	reward		0.93197626	
	std		1.07	
	value_loss		3.81	

	time/			
	fps		104	
	iterations		9400	
	time_elapsed		450	
	total_timesteps		47000	
	train/			
	entropy_loss		-43.1	
	explained_variance		-0.0773	
	learning_rate		0.0007	
	n_updates		9399	
	policy_loss		129	
	reward		0.41273853	
	std		1.07	
	value_loss		9.06	

	time/			
	fps		104	
	iterations		9500	
	time_elapsed		455	
	total_timesteps		47500	
	train/			
	entropy_loss		-43.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		9499	
	policy_loss		141	
	reward		0.08912667	
	std		1.07	
	value_loss		13.2	

	time/			
	fps		104	
	iterations		9600	
	time_elapsed		460	
	total_timesteps		48000	
	train/			
	entropy_loss		-43.1	
	explained_variance		-1.19e-07	
	learning_rate		0.0007	

	n_updates		9599	
	policy_loss		-89.5	
	reward		0.13666596	
	std		1.07	
	value_loss		5.14	

	time/			
	fps		104	
	iterations		9700	
	time_elapsed		465	
	total_timesteps		48500	
	train/			
	entropy_loss		-43.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		9699	
	policy_loss		94.8	
	reward		0.38808212	
	std		1.07	
	value_loss		7.65	

	time/			
	fps		104	
	iterations		9800	
	time_elapsed		469	
	total_timesteps		49000	
	train/			
	entropy_loss		-43.1	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		9799	
	policy_loss		324	
	reward		6.8032737	
	std		1.07	
	value_loss		59.4	

	time/			
	fps		104	
	iterations		9900	
	time_elapsed		474	
	total_timesteps		49500	
	train/			
	entropy_loss		-43.2	
	explained_variance		0	
	learning_rate		0.0007	

	n_updates		9899	
	policy_loss		30.1	
	reward		0.0035519307	
	std		1.07	
	value_loss		0.718	

---

	time/			
	fps		104	
	iterations		10000	
	time_elapsed		479	
	total_timesteps		50000	
	train/			
	entropy_loss		-43.2	
	explained_variance		0	
	learning_rate		0.0007	
	n_updates		9999	
	policy_loss		28.1	
	reward		-0.43205363	
	std		1.07	
	value_loss		1.01	

---

```
[9]: trained_a2c.save(TRAINED_MODEL_DIR + "/agent_a2c") if if_using_a2c else None
```

## 1.0.2 Agent 2: DDPG

```
[10]: agent = DRLAgent(env = env_train)
model_ddpg = agent.get_model("ddpg")

if if_using_ddpg:
    # set up logger
    tmp_path = RESULTS_DIR + '/ddpg'
    new_logger_ddpg = configure(tmp_path, ["stdout", "csv", "tensorboard"])
    # Set new logger
    model_ddpg.set_logger(new_logger_ddpg)
```

```
{'batch_size': 128, 'buffer_size': 50000, 'learning_rate': 0.001}
```

Using cuda device

Logging to results/ddpg

```
[11]: trained_ddpg = agent.train_model(model=model_ddpg,
                                       tb_log_name='ddpg',
                                       total_timesteps=50000) if if_using_ddpg else None
```

day: 2892, episode: 20

begin\_total\_asset: 1000000.00

end\_total\_asset: 3990512.90

total\_reward: 2990512.90

total\_cost: 5286.30

total\_trades: 55707

Sharpe: 0.800

```
=====
-----
| time/          |          |
|   episodes     | 4        |
|   fps          | 77       |
|   time_elapsed | 148      |
|   total_timesteps | 11572   |
| train/         |          |
|   actor_loss   | -49.5    |
|   critic_loss  | 135      |
|   learning_rate | 0.001    |
|   n_updates    | 8679     |
|   reward       | 4.3509665 |
-----
```

```
-----
| time/          |          |
|   episodes     | 8        |
|   fps          | 73       |
|   time_elapsed | 313      |
|   total_timesteps | 23144   |
| train/         |          |
|   actor_loss   | -29.5    |
|   critic_loss  | 11.2     |
|   learning_rate | 0.001    |
|   n_updates    | 20251    |
|   reward       | 4.3509665 |
-----
```

day: 2892, episode: 30

begin\_total\_asset: 1000000.00

end\_total\_asset: 4038764.75

total\_reward: 3038764.75

total\_cost: 999.00

total\_trades: 46245

Sharpe: 0.784

```
=====
-----
| time/          |          |
|   episodes     | 12       |
|   fps          | 67       |
|   time_elapsed | 515      |
|   total_timesteps | 34716   |
| train/         |          |
|   actor_loss   | -20.2    |
|   critic_loss  | 2.29     |
-----
```

	learning_rate	0.001	
	n_updates	31823	
	reward	4.3509665	
-----			
	time/		
	episodes	16	
	fps	67	
	time_elapsed	690	
	total_timesteps	46288	
	train/		
	actor_loss	-15.6	
	critic_loss	1.53	
	learning_rate	0.001	
	n_updates	43395	
	reward	4.3509665	
-----			

```
[12]: trained_ddpg.save(TRAINED_MODEL_DIR + "/agent_ddpg") if if_using_ddpg else None
```

### 1.0.3 Agent 3: PPO

```
[13]: agent = DRLAgent(env = env_train)
PPO_PARAMS = {
    "n_steps": 2048,
    "ent_coef": 0.01,
    "learning_rate": 0.00025,
    "batch_size": 128,
}
model_ppo = agent.get_model("ppo", model_kwargs = PPO_PARAMS)

if if_using_ppo:
    # set up logger
    tmp_path = RESULTS_DIR + '/ppo'
    new_logger_ppo = configure(tmp_path, ["stdout", "csv", "tensorboard"])
    # Set new logger
    model_ppo.set_logger(new_logger_ppo)
```

```
{'n_steps': 2048, 'ent_coef': 0.01, 'learning_rate': 0.00025, 'batch_size': 128}
Using cuda device
Logging to results/ppo
```

```
[14]: trained_ppo = agent.train_model(model=model_ppo,
                                     tb_log_name='ppo',
                                     total_timesteps=200000) if if_using_ppo else None
```

	time/		
--	-------	--	--

	fps	97	
	iterations	1	
	time_elapsed	21	
	total_timesteps	2048	
	train/		
	reward	0.020233056	

	time/		
	fps	100	
	iterations	2	
	time_elapsed	40	
	total_timesteps	4096	
	train/		
	approx_kl	0.016614027	
	clip_fraction	0.214	
	clip_range	0.2	
	entropy_loss	-41.2	
	explained_variance	-0.00975	
	learning_rate	0.00025	
	loss	3.58	
	n_updates	10	
	policy_gradient_loss	-0.025	
	reward	1.1401734	
	std	1	
	value_loss	9.79	

day: 2892, episode: 40  
 begin\_total\_asset: 1000000.00  
 end\_total\_asset: 3830414.75  
 total\_reward: 2830414.75  
 total\_cost: 337624.52  
 total\_trades: 80519  
 Sharpe: 0.689

	time/		
	fps	100	
	iterations	3	
	time_elapsed	61	
	total_timesteps	6144	
	train/		
	approx_kl	0.016332332	
	clip_fraction	0.182	
	clip_range	0.2	
	entropy_loss	-41.3	
	explained_variance	-0.00348	
	learning_rate	0.00025	

	loss	28.9	
	n_updates	20	
	policy_gradient_loss	-0.0158	
	reward	-1.5551344	
	std	1	
	value_loss	61.8	

---

	time/		
	fps	97	
	iterations	4	
	time_elapsed	83	
	total_timesteps	8192	
	train/		
	approx_kl	0.008956104	
	clip_fraction	0.0753	
	clip_range	0.2	
	entropy_loss	-41.3	
	explained_variance	0.00582	
	learning_rate	0.00025	
	loss	141	
	n_updates	30	
	policy_gradient_loss	-0.0146	
	reward	3.1444545	
	std	1.01	
	value_loss	237	

---

	time/		
	fps	99	
	iterations	5	
	time_elapsed	103	
	total_timesteps	10240	
	train/		
	approx_kl	0.014296137	
	clip_fraction	0.142	
	clip_range	0.2	
	entropy_loss	-41.3	
	explained_variance	-0.0516	
	learning_rate	0.00025	
	loss	10.5	
	n_updates	40	
	policy_gradient_loss	-0.0189	
	reward	2.9111667	
	std	1.01	
	value_loss	25	

---



time/		
fps	101	
iterations	6	
time_elapsed	120	
total_timesteps	12288	
train/		
approx_kl	0.028209247	
clip_fraction	0.3	
clip_range	0.2	
entropy_loss	-41.4	
explained_variance	-0.00886	
learning_rate	0.00025	
loss	36.7	
n_updates	50	
policy_gradient_loss	-0.0111	
reward	2.1514475	
std	1.01	
value_loss	66.4	

---

time/		
fps	103	
iterations	7	
time_elapsed	137	
total_timesteps	14336	
train/		
approx_kl	0.016963318	
clip_fraction	0.193	
clip_range	0.2	
entropy_loss	-41.4	
explained_variance	0.00439	
learning_rate	0.00025	
loss	26	
n_updates	60	
policy_gradient_loss	-0.0164	
reward	1.7487807	
std	1.01	
value_loss	117	

---

time/		
fps	104	
iterations	8	
time_elapsed	157	
total_timesteps	16384	
train/		
approx_kl	0.017198525	
clip_fraction	0.151	

clip_range	0.2	
entropy_loss	-41.5	
explained_variance	-0.0177	
learning_rate	0.00025	
loss	15.4	
n_updates	70	
policy_gradient_loss	-0.0141	
reward	1.2313813	
std	1.01	
value_loss	31.2	

time/		
fps	103	
iterations	9	
time_elapsed	178	
total_timesteps	18432	
train/		
approx_kl	0.015560304	
clip_fraction	0.182	
clip_range	0.2	
entropy_loss	-41.5	
explained_variance	-0.00549	
learning_rate	0.00025	
loss	24.7	
n_updates	80	
policy_gradient_loss	-0.0178	
reward	-0.05745077	
std	1.01	
value_loss	69.9	

time/		
fps	102	
iterations	10	
time_elapsed	199	
total_timesteps	20480	
train/		
approx_kl	0.023841653	
clip_fraction	0.241	
clip_range	0.2	
entropy_loss	-41.6	
explained_variance	0.00486	
learning_rate	0.00025	
loss	24.8	
n_updates	90	
policy_gradient_loss	-0.0139	
reward	1.0090605	

	std	1.02	
	value_loss	71.9	

	time/		
	fps	102	
	iterations	11	
	time_elapsed	220	
	total_timesteps	22528	
	train/		
	approx_kl	0.017691765	
	clip_fraction	0.196	
	clip_range	0.2	
	entropy_loss	-41.6	
	explained_variance	0.00431	
	learning_rate	0.00025	
	loss	23.3	
	n_updates	100	
	policy_gradient_loss	-0.0111	
	reward	-2.7392414	
	std	1.02	
	value_loss	50.5	

	time/		
	fps	101	
	iterations	12	
	time_elapsed	242	
	total_timesteps	24576	
	train/		
	approx_kl	0.020781098	
	clip_fraction	0.212	
	clip_range	0.2	
	entropy_loss	-41.7	
	explained_variance	0.00696	
	learning_rate	0.00025	
	loss	8.54	
	n_updates	110	
	policy_gradient_loss	-0.0233	
	reward	-0.25397715	
	std	1.02	
	value_loss	25.1	

	time/		
	fps	100	
	iterations	13	
	time_elapsed	265	

total_timesteps	26624
train/	
approx_kl	0.020514071
clip_fraction	0.199
clip_range	0.2
entropy_loss	-41.7
explained_variance	-0.00208
learning_rate	0.00025
loss	76.2
n_updates	120
policy_gradient_loss	-0.0116
reward	0.31476477
std	1.02
value_loss	108

time/	
fps	100
iterations	14
time_elapsed	284
total_timesteps	28672
train/	
approx_kl	0.019134384
clip_fraction	0.192
clip_range	0.2
entropy_loss	-41.8
explained_variance	0.00458
learning_rate	0.00025
loss	58.7
n_updates	130
policy_gradient_loss	-0.0144
reward	0.040682994
std	1.02
value_loss	94

time/	
fps	101
iterations	15
time_elapsed	303
total_timesteps	30720
train/	
approx_kl	0.017354943
clip_fraction	0.186
clip_range	0.2
entropy_loss	-41.8
explained_variance	-0.0296
learning_rate	0.00025

	loss		9.52	
	n_updates		140	
	policy_gradient_loss		-0.0132	
	reward		3.7676249	
	std		1.02	
	value_loss		23.7	

	time/			
	fps		101	
	iterations		16	
	time_elapsed		322	
	total_timesteps		32768	
	train/			
	approx_kl		0.020209845	
	clip_fraction		0.228	
	clip_range		0.2	
	entropy_loss		-41.9	
	explained_variance		-0.00256	
	learning_rate		0.00025	
	loss		15.2	
	n_updates		150	
	policy_gradient_loss		-0.0217	
	reward		-0.42414075	
	std		1.03	
	value_loss		43	

day: 2892, episode: 50  
 begin\_total\_asset: 1000000.00  
 end\_total\_asset: 5046854.87  
 total\_reward: 4046854.87  
 total\_cost: 337720.10  
 total\_trades: 80151  
 Sharpe: 0.870

	time/			
	fps		101	
	iterations		17	
	time_elapsed		341	
	total_timesteps		34816	
	train/			
	approx_kl		0.021866392	
	clip_fraction		0.228	
	clip_range		0.2	
	entropy_loss		-41.9	
	explained_variance		-0.00338	
	learning_rate		0.00025	

	loss		60.2	
	n_updates		160	
	policy_gradient_loss		-0.02	
	reward		0.804032	
	std		1.03	
	value_loss		75.7	

	time/			
	fps		102	
	iterations		18	
	time_elapsed		360	
	total_timesteps		36864	
	train/			
	approx_kl		0.021577742	
	clip_fraction		0.185	
	clip_range		0.2	
	entropy_loss		-42	
	explained_variance		-0.0187	
	learning_rate		0.00025	
	loss		39	
	n_updates		170	
	policy_gradient_loss		-0.0157	
	reward		-1.0861422	
	std		1.03	
	value_loss		132	

	time/			
	fps		101	
	iterations		19	
	time_elapsed		382	
	total_timesteps		38912	
	train/			
	approx_kl		0.03598381	
	clip_fraction		0.272	
	clip_range		0.2	
	entropy_loss		-42	
	explained_variance		-0.0245	
	learning_rate		0.00025	
	loss		9.39	
	n_updates		180	
	policy_gradient_loss		-0.0186	
	reward		-0.51048476	
	std		1.03	
	value_loss		20.7	

time/		
fps	101	
iterations	20	
time_elapsed	404	
total_timesteps	40960	
train/		
approx_kl	0.022985812	
clip_fraction	0.21	
clip_range	0.2	
entropy_loss	-42.1	
explained_variance	-0.000124	
learning_rate	0.00025	
loss	56.9	
n_updates	190	
policy_gradient_loss	-0.0203	
reward	-0.17416683	
std	1.04	
value_loss	87.2	

---

time/		
fps	101	
iterations	21	
time_elapsed	424	
total_timesteps	43008	
train/		
approx_kl	0.025186807	
clip_fraction	0.27	
clip_range	0.2	
entropy_loss	-42.2	
explained_variance	0.00742	
learning_rate	0.00025	
loss	16.1	
n_updates	200	
policy_gradient_loss	-0.016	
reward	-7.5072427	
std	1.04	
value_loss	53.8	

---

time/		
fps	101	
iterations	22	
time_elapsed	444	
total_timesteps	45056	
train/		
approx_kl	0.029560171	
clip_fraction	0.272	

clip_range	0.2	
entropy_loss	-42.2	
explained_variance	-0.00341	
learning_rate	0.00025	
loss	14	
n_updates	210	
policy_gradient_loss	-0.0151	
reward	2.710092	
std	1.04	
value_loss	29.2	

time/		
fps	101	
iterations	23	
time_elapsed	463	
total_timesteps	47104	
train/		
approx_kl	0.03276337	
clip_fraction	0.216	
clip_range	0.2	
entropy_loss	-42.2	
explained_variance	0.005	
learning_rate	0.00025	
loss	42.4	
n_updates	220	
policy_gradient_loss	-0.0135	
reward	-0.15345418	
std	1.04	
value_loss	104	

time/		
fps	101	
iterations	24	
time_elapsed	483	
total_timesteps	49152	
train/		
approx_kl	0.03130112	
clip_fraction	0.288	
clip_range	0.2	
entropy_loss	-42.3	
explained_variance	0.00448	
learning_rate	0.00025	
loss	60.9	
n_updates	230	
policy_gradient_loss	-0.0167	
reward	7.858115	



std	1.04
value_loss	60

time/	
fps	102
iterations	25
time_elapsed	501
total_timesteps	51200
train/	
approx_kl	0.035679255
clip_fraction	0.294
clip_range	0.2
entropy_loss	-42.3
explained_variance	0.00272
learning_rate	0.00025
loss	67.9
n_updates	240
policy_gradient_loss	-0.0121
reward	-0.1196461
std	1.04
value_loss	124

time/	
fps	102
iterations	26
time_elapsed	521
total_timesteps	53248
train/	
approx_kl	0.02294702
clip_fraction	0.257
clip_range	0.2
entropy_loss	-42.4
explained_variance	-0.00835
learning_rate	0.00025
loss	16.7
n_updates	250
policy_gradient_loss	-0.0142
reward	0.79509485
std	1.05
value_loss	29.3

time/	
fps	102
iterations	27
time_elapsed	539

total_timesteps	55296
train/	
approx_kl	0.029297683
clip_fraction	0.258
clip_range	0.2
entropy_loss	-42.5
explained_variance	0.00436
learning_rate	0.00025
loss	55.4
n_updates	260
policy_gradient_loss	-0.0168
reward	-0.6547185
std	1.05
value_loss	120

time/	
fps	102
iterations	28
time_elapsed	557
total_timesteps	57344
train/	
approx_kl	0.030363379
clip_fraction	0.26
clip_range	0.2
entropy_loss	-42.6
explained_variance	0.00645
learning_rate	0.00025
loss	36.2
n_updates	270
policy_gradient_loss	-0.0142
reward	-0.365066
std	1.05
value_loss	76.9

time/	
fps	103
iterations	29
time_elapsed	575
total_timesteps	59392
train/	
approx_kl	0.027431753
clip_fraction	0.274
clip_range	0.2
entropy_loss	-42.6
explained_variance	0.0059
learning_rate	0.00025

	loss		14.7	
	n_updates		280	
	policy_gradient_loss		-0.0144	
	reward		4.1839476	
	std		1.05	
	value_loss		23.7	

	time/			
	fps		103	
	iterations		30	
	time_elapsed		592	
	total_timesteps		61440	
	train/			
	approx_kl		0.022260884	
	clip_fraction		0.214	
	clip_range		0.2	
	entropy_loss		-42.6	
	explained_variance		0.0014	
	learning_rate		0.00025	
	loss		15.9	
	n_updates		290	
	policy_gradient_loss		-0.0147	
	reward		0.4827492	
	std		1.05	
	value_loss		66.5	

	time/			
	fps		104	
	iterations		31	
	time_elapsed		609	
	total_timesteps		63488	
	train/			
	approx_kl		0.02043109	
	clip_fraction		0.192	
	clip_range		0.2	
	entropy_loss		-42.7	
	explained_variance		-0.00464	
	learning_rate		0.00025	
	loss		49.7	
	n_updates		300	
	policy_gradient_loss		-0.018	
	reward		5.625035	
	std		1.05	
	value_loss		96.8	

day: 2892, episode: 60

```

begin_total_asset: 1000000.00
end_total_asset: 4052945.55
total_reward: 3052945.55
total_cost: 313074.34
total_trades: 77783
Sharpe: 0.840

```

```
=====
```

-----		
time/		
fps	104	
iterations	32	
time_elapsed	627	
total_timesteps	65536	
train/		
approx_kl	0.029359937	
clip_fraction	0.278	
clip_range	0.2	
entropy_loss	-42.7	
explained_variance	0.00628	
learning_rate	0.00025	
loss	19.1	
n_updates	310	
policy_gradient_loss	-0.00768	
reward	-0.047850505	
std	1.06	
value_loss	37.3	

```
-----
```

```
-----
```

time/		
fps	104	
iterations	33	
time_elapsed	645	
total_timesteps	67584	
train/		
approx_kl	0.030758934	
clip_fraction	0.291	
clip_range	0.2	
entropy_loss	-42.8	
explained_variance	-0.0273	
learning_rate	0.00025	
loss	14.6	
n_updates	320	
policy_gradient_loss	-0.0218	
reward	-0.53193337	
std	1.06	
value_loss	54.2	

```
-----
```

time/		
fps	105	
iterations	34	
time_elapsed	662	
total_timesteps	69632	
train/		
approx_kl	0.028932806	
clip_fraction	0.29	
clip_range	0.2	
entropy_loss	-42.9	
explained_variance	0.00686	
learning_rate	0.00025	
loss	25.7	
n_updates	330	
policy_gradient_loss	-0.0112	
reward	1.1589198	
std	1.06	
value_loss	53.3	

---

time/		
fps	105	
iterations	35	
time_elapsed	680	
total_timesteps	71680	
train/		
approx_kl	0.027908195	
clip_fraction	0.293	
clip_range	0.2	
entropy_loss	-42.9	
explained_variance	0.00503	
learning_rate	0.00025	
loss	48	
n_updates	340	
policy_gradient_loss	-0.00996	
reward	0.8199292	
std	1.06	
value_loss	215	

---

time/		
fps	105	
iterations	36	
time_elapsed	698	
total_timesteps	73728	
train/		
approx_kl	0.040287144	
clip_fraction	0.363	

clip_range	0.2	
entropy_loss	-43	
explained_variance	-0.0248	
learning_rate	0.00025	
loss	8.48	
n_updates	350	
policy_gradient_loss	-0.0112	
reward	-5.082193	
std	1.06	
value_loss	17.3	

time/		
fps	105	
iterations	37	
time_elapsed	716	
total_timesteps	75776	
train/		
approx_kl	0.03675144	
clip_fraction	0.303	
clip_range	0.2	
entropy_loss	-43	
explained_variance	0.00886	
learning_rate	0.00025	
loss	40.1	
n_updates	360	
policy_gradient_loss	-0.0145	
reward	-0.51240253	
std	1.07	
value_loss	63.7	

time/		
fps	105	
iterations	38	
time_elapsed	734	
total_timesteps	77824	
train/		
approx_kl	0.033948477	
clip_fraction	0.329	
clip_range	0.2	
entropy_loss	-43.1	
explained_variance	-0.00215	
learning_rate	0.00025	
loss	104	
n_updates	370	
policy_gradient_loss	-0.00612	
reward	-7.0982575	

std	1.07
value_loss	128

time/	
fps	106
iterations	39
time_elapsed	751
total_timesteps	79872
train/	
approx_kl	0.022807235
clip_fraction	0.283
clip_range	0.2
entropy_loss	-43.2
explained_variance	0.0271
learning_rate	0.00025
loss	21.2
n_updates	380
policy_gradient_loss	-0.00597
reward	-3.5229821
std	1.07
value_loss	39.6

time/	
fps	106
iterations	40
time_elapsed	769
total_timesteps	81920
train/	
approx_kl	0.03216366
clip_fraction	0.317
clip_range	0.2
entropy_loss	-43.2
explained_variance	0.00291
learning_rate	0.00025
loss	10.3
n_updates	390
policy_gradient_loss	-0.0118
reward	-1.0320076
std	1.08
value_loss	77.7

time/	
fps	106
iterations	41
time_elapsed	787

total_timesteps	83968
train/	
approx_kl	0.036064602
clip_fraction	0.329
clip_range	0.2
entropy_loss	-43.3
explained_variance	0.0112
learning_rate	0.00025
loss	24.6
n_updates	400
policy_gradient_loss	-0.00544
reward	0.38228673
std	1.08
value_loss	120

time/	
fps	106
iterations	42
time_elapsed	804
total_timesteps	86016
train/	
approx_kl	0.027661435
clip_fraction	0.283
clip_range	0.2
entropy_loss	-43.4
explained_variance	-0.0104
learning_rate	0.00025
loss	20
n_updates	410
policy_gradient_loss	-0.00657
reward	-0.4453181
std	1.08
value_loss	58.6

time/	
fps	107
iterations	43
time_elapsed	822
total_timesteps	88064
train/	
approx_kl	0.027768198
clip_fraction	0.294
clip_range	0.2
entropy_loss	-43.4
explained_variance	0.0153
learning_rate	0.00025



	loss		9.35	
	n_updates		420	
	policy_gradient_loss		-0.017	
	reward		-0.5950569	
	std		1.08	
	value_loss		19.9	

---

	time/			
	fps		107	
	iterations		44	
	time_elapsed		839	
	total_timesteps		90112	
	train/			
	approx_kl		0.030285668	
	clip_fraction		0.292	
	clip_range		0.2	
	entropy_loss		-43.5	
	explained_variance		0.0032	
	learning_rate		0.00025	
	loss		59	
	n_updates		430	
	policy_gradient_loss		-0.00738	
	reward		0.2528765	
	std		1.09	
	value_loss		93.9	

---

	time/			
	fps		107	
	iterations		45	
	time_elapsed		857	
	total_timesteps		92160	
	train/			
	approx_kl		0.029212791	
	clip_fraction		0.257	
	clip_range		0.2	
	entropy_loss		-43.5	
	explained_variance		-0.00514	
	learning_rate		0.00025	
	loss		64.1	
	n_updates		440	
	policy_gradient_loss		-0.0161	
	reward		0.2864157	
	std		1.09	
	value_loss		102	

day: 2892, episode: 70

```

begin_total_asset: 1000000.00
end_total_asset: 6223713.95
total_reward: 5223713.95
total_cost: 325461.83
total_trades: 78189
Sharpe: 0.999

```

```
=====
```

-----		
time/		
fps	107	
iterations	46	
time_elapsed	874	
total_timesteps	94208	
train/		
approx_kl	0.04643644	
clip_fraction	0.334	
clip_range	0.2	
entropy_loss	-43.6	
explained_variance	-0.00391	
learning_rate	0.00025	
loss	23.2	
n_updates	450	
policy_gradient_loss	-0.00864	
reward	-5.471007	
std	1.09	
value_loss	44.9	
-----		

-----		
time/		
fps	107	
iterations	47	
time_elapsed	892	
total_timesteps	96256	
train/		
approx_kl	0.02952569	
clip_fraction	0.288	
clip_range	0.2	
entropy_loss	-43.7	
explained_variance	-0.0036	
learning_rate	0.00025	
loss	112	
n_updates	460	
policy_gradient_loss	-0.0135	
reward	1.1923963	
std	1.09	
value_loss	133	
-----		

```
-----
```

time/		
fps	108	
iterations	48	
time_elapsed	909	
total_timesteps	98304	
train/		
approx_kl	0.026492959	
clip_fraction	0.267	
clip_range	0.2	
entropy_loss	-43.7	
explained_variance	-0.00489	
learning_rate	0.00025	
loss	33.9	
n_updates	470	
policy_gradient_loss	-0.0102	
reward	12.741209	
std	1.09	
value_loss	86.3	

---

time/		
fps	108	
iterations	49	
time_elapsed	927	
total_timesteps	100352	
train/		
approx_kl	0.029927118	
clip_fraction	0.277	
clip_range	0.2	
entropy_loss	-43.8	
explained_variance	-0.00295	
learning_rate	0.00025	
loss	25.6	
n_updates	480	
policy_gradient_loss	-0.00958	
reward	-1.3748448	
std	1.1	
value_loss	136	

---

time/		
fps	108	
iterations	50	
time_elapsed	944	
total_timesteps	102400	
train/		
approx_kl	0.03208095	
clip_fraction	0.293	

clip_range	0.2	
entropy_loss	-43.9	
explained_variance	0.00985	
learning_rate	0.00025	
loss	29.8	
n_updates	490	
policy_gradient_loss	-0.0194	
reward	0.42778212	
std	1.1	
value_loss	50.1	

time/		
fps	108	
iterations	51	
time_elapsed	962	
total_timesteps	104448	
train/		
approx_kl	0.026934743	
clip_fraction	0.275	
clip_range	0.2	
entropy_loss	-43.9	
explained_variance	0.0101	
learning_rate	0.00025	
loss	47	
n_updates	500	
policy_gradient_loss	-0.0133	
reward	-0.12136642	
std	1.1	
value_loss	233	

time/		
fps	108	
iterations	52	
time_elapsed	979	
total_timesteps	106496	
train/		
approx_kl	0.038600534	
clip_fraction	0.356	
clip_range	0.2	
entropy_loss	-44	
explained_variance	0.00029	
learning_rate	0.00025	
loss	115	
n_updates	510	
policy_gradient_loss	-0.00986	
reward	-3.4217103	

	std	1.11	
	value_loss	170	
-----			
	time/		
	fps	108	
	iterations	53	
	time_elapsed	997	
	total_timesteps	108544	
	train/		
	approx_kl	0.036651038	
	clip_fraction	0.34	
	clip_range	0.2	
	entropy_loss	-44.1	
	explained_variance	0.0288	
	learning_rate	0.00025	
	loss	16.7	
	n_updates	520	
	policy_gradient_loss	-0.00926	
	reward	3.7280262	
	std	1.11	
	value_loss	41.8	
-----			
	time/		
	fps	108	
	iterations	54	
	time_elapsed	1015	
	total_timesteps	110592	
	train/		
	approx_kl	0.028470214	
	clip_fraction	0.28	
	clip_range	0.2	
	entropy_loss	-44.2	
	explained_variance	0.00968	
	learning_rate	0.00025	
	loss	33.3	
	n_updates	530	
	policy_gradient_loss	-0.00552	
	reward	1.2410479	
	std	1.11	
	value_loss	127	
-----			
	time/		
	fps	109	
	iterations	55	
	time_elapsed	1032	

total_timesteps	112640
train/	
approx_kl	0.026078358
clip_fraction	0.193
clip_range	0.2
entropy_loss	-44.3
explained_variance	0.0179
learning_rate	0.00025
loss	86.4
n_updates	540
policy_gradient_loss	-0.00869
reward	0.5561161
std	1.12
value_loss	168

time/	
fps	109
iterations	56
time_elapsed	1049
total_timesteps	114688
train/	
approx_kl	0.024531223
clip_fraction	0.217
clip_range	0.2
entropy_loss	-44.3
explained_variance	0.0175
learning_rate	0.00025
loss	17.4
n_updates	550
policy_gradient_loss	-0.00437
reward	1.9761248
std	1.12
value_loss	50.9

time/	
fps	109
iterations	57
time_elapsed	1067
total_timesteps	116736
train/	
approx_kl	0.029285105
clip_fraction	0.278
clip_range	0.2
entropy_loss	-44.3
explained_variance	0.0076
learning_rate	0.00025

	loss		38.1	
	n_updates		560	
	policy_gradient_loss		-0.0108	
	reward		0.31480572	
	std		1.12	
	value_loss		100	

	time/			
	fps		109	
	iterations		58	
	time_elapsed		1084	
	total_timesteps		118784	
	train/			
	approx_kl		0.027524665	
	clip_fraction		0.238	
	clip_range		0.2	
	entropy_loss		-44.4	
	explained_variance		0.014	
	learning_rate		0.00025	
	loss		61.6	
	n_updates		570	
	policy_gradient_loss		-0.00468	
	reward		0.67348945	
	std		1.12	
	value_loss		125	

	time/			
	fps		109	
	iterations		59	
	time_elapsed		1102	
	total_timesteps		120832	
	train/			
	approx_kl		0.035444025	
	clip_fraction		0.335	
	clip_range		0.2	
	entropy_loss		-44.4	
	explained_variance		-0.00158	
	learning_rate		0.00025	
	loss		70.3	
	n_updates		580	
	policy_gradient_loss		-0.00113	
	reward		-1.8806355	
	std		1.12	
	value_loss		155	

day: 2892, episode: 80

```

begin_total_asset: 1000000.00
end_total_asset: 4367320.71
total_reward: 3367320.71
total_cost: 299417.35
total_trades: 76503
Sharpe: 0.796

```

```
=====
```

-----		
time/		
fps	109	
iterations	60	
time_elapsed	1120	
total_timesteps	122880	
train/		
approx_kl	0.03247026	
clip_fraction	0.322	
clip_range	0.2	
entropy_loss	-44.5	
explained_variance	0.0697	
learning_rate	0.00025	
loss	16.7	
n_updates	590	
policy_gradient_loss	-0.0134	
reward	0.21105283	
std	1.12	
value_loss	30.7	
-----		

-----		
time/		
fps	109	
iterations	61	
time_elapsed	1138	
total_timesteps	124928	
train/		
approx_kl	0.025711142	
clip_fraction	0.279	
clip_range	0.2	
entropy_loss	-44.6	
explained_variance	0.0289	
learning_rate	0.00025	
loss	44.4	
n_updates	600	
policy_gradient_loss	-0.00982	
reward	0.86830705	
std	1.13	
value_loss	133	
-----		

```
-----
```



time/		
fps	109	
iterations	62	
time_elapsed	1157	
total_timesteps	126976	
train/		
approx_kl	0.03209849	
clip_fraction	0.28	
clip_range	0.2	
entropy_loss	-44.7	
explained_variance	0.0222	
learning_rate	0.00025	
loss	86.6	
n_updates	610	
policy_gradient_loss	-0.00477	
reward	6.3152614	
std	1.13	
value_loss	152	

---

time/		
fps	109	
iterations	63	
time_elapsed	1174	
total_timesteps	129024	
train/		
approx_kl	0.021212902	
clip_fraction	0.203	
clip_range	0.2	
entropy_loss	-44.7	
explained_variance	0.00386	
learning_rate	0.00025	
loss	18.6	
n_updates	620	
policy_gradient_loss	-0.00633	
reward	5.6451497	
std	1.13	
value_loss	39.5	

---

time/		
fps	109	
iterations	64	
time_elapsed	1192	
total_timesteps	131072	
train/		
approx_kl	0.028615076	
clip_fraction	0.25	

clip_range	0.2	
entropy_loss	-44.8	
explained_variance	0.0273	
learning_rate	0.00025	
loss	63.5	
n_updates	630	
policy_gradient_loss	-0.0136	
reward	-0.22170344	
std	1.14	
value_loss	100	

time/		
fps	110	
iterations	65	
time_elapsed	1209	
total_timesteps	133120	
train/		
approx_kl	0.025787722	
clip_fraction	0.241	
clip_range	0.2	
entropy_loss	-44.9	
explained_variance	0.0259	
learning_rate	0.00025	
loss	61.7	
n_updates	640	
policy_gradient_loss	-0.0107	
reward	-1.4330577	
std	1.14	
value_loss	154	

time/		
fps	110	
iterations	66	
time_elapsed	1227	
total_timesteps	135168	
train/		
approx_kl	0.044934466	
clip_fraction	0.381	
clip_range	0.2	
entropy_loss	-45	
explained_variance	0.00543	
learning_rate	0.00025	
loss	73.7	
n_updates	650	
policy_gradient_loss	0.00588	
reward	2.7523956	

	std	1.14	
	value_loss	167	

	time/		
	fps	110	
	iterations	67	
	time_elapsed	1244	
	total_timesteps	137216	
	train/		
	approx_kl	0.042157225	
	clip_fraction	0.316	
	clip_range	0.2	
	entropy_loss	-45	
	explained_variance	0.137	
	learning_rate	0.00025	
	loss	13.6	
	n_updates	660	
	policy_gradient_loss	-0.0101	
	reward	-1.7714602	
	std	1.14	
	value_loss	29.4	

	time/		
	fps	110	
	iterations	68	
	time_elapsed	1262	
	total_timesteps	139264	
	train/		
	approx_kl	0.028857064	
	clip_fraction	0.285	
	clip_range	0.2	
	entropy_loss	-45.1	
	explained_variance	0.0473	
	learning_rate	0.00025	
	loss	39.5	
	n_updates	670	
	policy_gradient_loss	-0.00981	
	reward	0.10262209	
	std	1.15	
	value_loss	150	

	time/		
	fps	110	
	iterations	69	
	time_elapsed	1280	

total_timesteps	141312
train/	
approx_kl	0.035281073
clip_fraction	0.247
clip_range	0.2
entropy_loss	-45.1
explained_variance	0.0207
learning_rate	0.00025
loss	61.6
n_updates	680
policy_gradient_loss	-0.00538
reward	1.894777
std	1.15
value_loss	150

time/	
fps	110
iterations	70
time_elapsed	1297
total_timesteps	143360
train/	
approx_kl	0.04755744
clip_fraction	0.347
clip_range	0.2
entropy_loss	-45.1
explained_variance	0.116
learning_rate	0.00025
loss	25.7
n_updates	690
policy_gradient_loss	-0.000433
reward	1.0903006
std	1.15
value_loss	46.5

time/	
fps	110
iterations	71
time_elapsed	1315
total_timesteps	145408
train/	
approx_kl	0.024918608
clip_fraction	0.268
clip_range	0.2
entropy_loss	-45.2
explained_variance	-0.00579
learning_rate	0.00025

	loss		55.8	
	n_updates		700	
	policy_gradient_loss		-0.0138	
	reward		0.8389404	
	std		1.15	
	value_loss		154	

---

	time/			
	fps		110	
	iterations		72	
	time_elapsed		1333	
	total_timesteps		147456	
	train/			
	approx_kl		0.029028082	
	clip_fraction		0.234	
	clip_range		0.2	
	entropy_loss		-45.2	
	explained_variance		0.00158	
	learning_rate		0.00025	
	loss		126	
	n_updates		710	
	policy_gradient_loss		-0.00434	
	reward		-17.191538	
	std		1.15	
	value_loss		228	

---

	time/			
	fps		110	
	iterations		73	
	time_elapsed		1350	
	total_timesteps		149504	
	train/			
	approx_kl		0.022142101	
	clip_fraction		0.199	
	clip_range		0.2	
	entropy_loss		-45.3	
	explained_variance		0.0602	
	learning_rate		0.00025	
	loss		29.1	
	n_updates		720	
	policy_gradient_loss		-0.00667	
	reward		-0.85172176	
	std		1.15	
	value_loss		55.7	

day: 2892, episode: 90

```

begin_total_asset: 1000000.00
end_total_asset: 5952479.32
total_reward: 4952479.32
total_cost: 277112.61
total_trades: 75548
Sharpe: 0.932
=====

```

```

-----
| time/          |          |
|   fps          |   110    |
|   iterations    |   74     |
|   time_elapsed  |  1368    |
|   total_timesteps | 151552   |
| train/         |          |
|   approx_kl     | 0.027287846 |
|   clip_fraction | 0.283     |
|   clip_range    | 0.2       |
|   entropy_loss  | -45.3     |
|   explained_variance | 0.0856    |
|   learning_rate | 0.00025   |
|   loss          | 26.5      |
|   n_updates     | 730       |
|   policy_gradient_loss | -0.0113   |
|   reward        | -2.0646653 |
|   std           | 1.16      |
|   value_loss    | 50.2      |
-----

```

```

-----
| time/          |          |
|   fps          |   110    |
|   iterations    |   75     |
|   time_elapsed  |  1386    |
|   total_timesteps | 153600   |
| train/         |          |
|   approx_kl     | 0.03027108 |
|   clip_fraction | 0.263     |
|   clip_range    | 0.2       |
|   entropy_loss  | -45.3     |
|   explained_variance | 0.0266    |
|   learning_rate | 0.00025   |
|   loss          | 60.2      |
|   n_updates     | 740       |
|   policy_gradient_loss | -0.0117   |
|   reward        | 1.0992491 |
|   std           | 1.16      |
|   value_loss    | 166       |
-----

```

time/		
fps	110	
iterations	76	
time_elapsed	1407	
total_timesteps	155648	
train/		
approx_kl	0.024960503	
clip_fraction	0.263	
clip_range	0.2	
entropy_loss	-45.3	
explained_variance	0.0344	
learning_rate	0.00025	
loss	65.7	
n_updates	750	
policy_gradient_loss	-0.0116	
reward	-2.112159	
std	1.16	
value_loss	146	

---

time/		
fps	110	
iterations	77	
time_elapsed	1427	
total_timesteps	157696	
train/		
approx_kl	0.04468357	
clip_fraction	0.414	
clip_range	0.2	
entropy_loss	-45.4	
explained_variance	0.0405	
learning_rate	0.00025	
loss	14.3	
n_updates	760	
policy_gradient_loss	0.0047	
reward	0.8548301	
std	1.16	
value_loss	28.2	

---

time/		
fps	110	
iterations	78	
time_elapsed	1447	
total_timesteps	159744	
train/		
approx_kl	0.021239018	
clip_fraction	0.222	

clip_range	0.2	
entropy_loss	-45.5	
explained_variance	0.0483	
learning_rate	0.00025	
loss	43.2	
n_updates	770	
policy_gradient_loss	-0.00619	
reward	2.9244356	
std	1.16	
value_loss	99.4	

time/		
fps	110	
iterations	79	
time_elapsed	1467	
total_timesteps	161792	
train/		
approx_kl	0.021285415	
clip_fraction	0.277	
clip_range	0.2	
entropy_loss	-45.5	
explained_variance	0.0354	
learning_rate	0.00025	
loss	73.4	
n_updates	780	
policy_gradient_loss	-0.0058	
reward	-4.517798	
std	1.17	
value_loss	152	

time/		
fps	110	
iterations	80	
time_elapsed	1488	
total_timesteps	163840	
train/		
approx_kl	0.035121735	
clip_fraction	0.249	
clip_range	0.2	
entropy_loss	-45.6	
explained_variance	0.057	
learning_rate	0.00025	
loss	34.5	
n_updates	790	
policy_gradient_loss	-0.00231	
reward	-0.4839865	



	std	1.17	
	value_loss	63.2	

	time/		
	fps	110	
	iterations	81	
	time_elapsed	1507	
	total_timesteps	165888	
	train/		
	approx_kl	0.035313547	
	clip_fraction	0.305	
	clip_range	0.2	
	entropy_loss	-45.6	
	explained_variance	0.00167	
	learning_rate	0.00025	
	loss	56.4	
	n_updates	800	
	policy_gradient_loss	-0.00533	
	reward	-0.49533653	
	std	1.17	
	value_loss	130	

	time/		
	fps	109	
	iterations	82	
	time_elapsed	1528	
	total_timesteps	167936	
	train/		
	approx_kl	0.03232641	
	clip_fraction	0.232	
	clip_range	0.2	
	entropy_loss	-45.7	
	explained_variance	0.0372	
	learning_rate	0.00025	
	loss	60.4	
	n_updates	810	
	policy_gradient_loss	-0.00415	
	reward	0.37259644	
	std	1.17	
	value_loss	98.3	

	time/		
	fps	109	
	iterations	83	
	time_elapsed	1548	

	total_timesteps		169984	
	train/			
	approx_kl		0.03147866	
	clip_fraction		0.336	
	clip_range		0.2	
	entropy_loss		-45.7	
	explained_variance		0.0301	
	learning_rate		0.00025	
	loss		45.9	
	n_updates		820	
	policy_gradient_loss		-0.00168	
	reward		-0.12214053	
	std		1.17	
	value_loss		183	

	time/			
	fps		109	
	iterations		84	
	time_elapsed		1568	
	total_timesteps		172032	
	train/			
	approx_kl		0.019143187	
	clip_fraction		0.212	
	clip_range		0.2	
	entropy_loss		-45.7	
	explained_variance		0.105	
	learning_rate		0.00025	
	loss		19.5	
	n_updates		830	
	policy_gradient_loss		-0.00581	
	reward		0.55154085	
	std		1.17	
	value_loss		29.1	

	time/			
	fps		109	
	iterations		85	
	time_elapsed		1588	
	total_timesteps		174080	
	train/			
	approx_kl		0.030926457	
	clip_fraction		0.266	
	clip_range		0.2	
	entropy_loss		-45.8	
	explained_variance		0.0515	
	learning_rate		0.00025	

	loss		20	
	n_updates		840	
	policy_gradient_loss		0.00125	
	reward		0.014717502	
	std		1.18	
	value_loss		111	

	time/			
	fps		109	
	iterations		86	
	time_elapsed		1608	
	total_timesteps		176128	
	train/			
	approx_kl		0.026563242	
	clip_fraction		0.346	
	clip_range		0.2	
	entropy_loss		-45.9	
	explained_variance		0.0336	
	learning_rate		0.00025	
	loss		54.4	
	n_updates		850	
	policy_gradient_loss		0.00268	
	reward		-1.2039502	
	std		1.18	
	value_loss		130	

	time/			
	fps		109	
	iterations		87	
	time_elapsed		1628	
	total_timesteps		178176	
	train/			
	approx_kl		0.029672872	
	clip_fraction		0.262	
	clip_range		0.2	
	entropy_loss		-45.9	
	explained_variance		0.0423	
	learning_rate		0.00025	
	loss		17.8	
	n_updates		860	
	policy_gradient_loss		-0.00304	
	reward		2.139257	
	std		1.18	
	value_loss		43	

day: 2892, episode: 100

```

begin_total_asset: 1000000.00
end_total_asset: 5440252.47
total_reward: 4440252.47
total_cost: 228788.22
total_trades: 72380
Sharpe: 0.917

```

```
=====
```

-----		
time/		
fps	109	
iterations	88	
time_elapsed	1648	
total_timesteps	180224	
train/		
approx_kl	0.023213124	
clip_fraction	0.235	
clip_range	0.2	
entropy_loss	-46	
explained_variance	0.0638	
learning_rate	0.00025	
loss	16.1	
n_updates	870	
policy_gradient_loss	-0.00368	
reward	-1.7594525	
std	1.18	
value_loss	124	
-----		

-----		
time/		
fps	109	
iterations	89	
time_elapsed	1668	
total_timesteps	182272	
train/		
approx_kl	0.034683395	
clip_fraction	0.3	
clip_range	0.2	
entropy_loss	-46	
explained_variance	-0.0018	
learning_rate	0.00025	
loss	45.6	
n_updates	880	
policy_gradient_loss	-0.000302	
reward	-0.09267654	
std	1.19	
value_loss	158	
-----		
-----		

time/		
fps	109	
iterations	90	
time_elapsed	1689	
total_timesteps	184320	
train/		
approx_kl	0.020717867	
clip_fraction	0.244	
clip_range	0.2	
entropy_loss	-46.1	
explained_variance	0.0586	
learning_rate	0.00025	
loss	70.1	
n_updates	890	
policy_gradient_loss	-0.00456	
reward	-0.87722206	
std	1.19	
value_loss	121	

---

time/		
fps	108	
iterations	91	
time_elapsed	1711	
total_timesteps	186368	
train/		
approx_kl	0.05710158	
clip_fraction	0.371	
clip_range	0.2	
entropy_loss	-46.2	
explained_variance	0.043	
learning_rate	0.00025	
loss	14.8	
n_updates	900	
policy_gradient_loss	0.0046	
reward	-1.1827924	
std	1.19	
value_loss	21.6	

---

time/		
fps	108	
iterations	92	
time_elapsed	1731	
total_timesteps	188416	
train/		
approx_kl	0.028326593	
clip_fraction	0.262	

clip_range	0.2	
entropy_loss	-46.3	
explained_variance	0.0249	
learning_rate	0.00025	
loss	32.1	
n_updates	910	
policy_gradient_loss	-0.00309	
reward	-1.9556996	
std	1.2	
value_loss	97.7	

time/		
fps	108	
iterations	93	
time_elapsed	1752	
total_timesteps	190464	
train/		
approx_kl	0.035645988	
clip_fraction	0.328	
clip_range	0.2	
entropy_loss	-46.3	
explained_variance	0.0119	
learning_rate	0.00025	
loss	45.2	
n_updates	920	
policy_gradient_loss	-0.00524	
reward	-0.11332548	
std	1.2	
value_loss	148	

time/		
fps	108	
iterations	94	
time_elapsed	1772	
total_timesteps	192512	
train/		
approx_kl	0.03274678	
clip_fraction	0.292	
clip_range	0.2	
entropy_loss	-46.3	
explained_variance	0.0765	
learning_rate	0.00025	
loss	7.87	
n_updates	930	
policy_gradient_loss	-0.0026	
reward	-0.008197777	

std	1.2
value_loss	31.9

time/	
fps	108
iterations	95
time_elapsed	1792
total_timesteps	194560
train/	
approx_kl	0.022962091
clip_fraction	0.267
clip_range	0.2
entropy_loss	-46.4
explained_variance	0.0302
learning_rate	0.00025
loss	50
n_updates	940
policy_gradient_loss	-0.00797
reward	-0.55041814
std	1.2
value_loss	119

time/	
fps	108
iterations	96
time_elapsed	1812
total_timesteps	196608
train/	
approx_kl	0.030306041
clip_fraction	0.274
clip_range	0.2
entropy_loss	-46.4
explained_variance	-0.0133
learning_rate	0.00025
loss	67.8
n_updates	950
policy_gradient_loss	-0.000764
reward	4.9244113
std	1.2
value_loss	206

time/	
fps	108
iterations	97
time_elapsed	1832

	total_timesteps		198656	
	train/			
	approx_kl		0.028664932	
	clip_fraction		0.267	
	clip_range		0.2	
	entropy_loss		-46.4	
	explained_variance		0.151	
	learning_rate		0.00025	
	loss		23.2	
	n_updates		960	
	policy_gradient_loss		0.0019	
	reward		0.21847676	
	std		1.2	
	value_loss		48.3	

	time/			
	fps		108	
	iterations		98	
	time_elapsed		1853	
	total_timesteps		200704	
	train/			
	approx_kl		0.027408203	
	clip_fraction		0.277	
	clip_range		0.2	
	entropy_loss		-46.5	
	explained_variance		0.00158	
	learning_rate		0.00025	
	loss		26.9	
	n_updates		970	
	policy_gradient_loss		-0.00807	
	reward		0.6061848	
	std		1.2	
	value_loss		90.7	

```
[15]: trained_ppo.save(TRAINED_MODEL_DIR + "/agent_ppo") if if_using_ppo else None
```

#### 1.0.4 Agent 4: SAC

```
[16]: agent = DRLAgent(env = env_train)
SAC_PARAMS = {
    "batch_size": 128,
    "buffer_size": 100000,
    "learning_rate": 0.0001,
    "learning_starts": 100,
    "ent_coef": "auto_0.1",
```



```

}

model_sac = agent.get_model("sac",model_kwargs = SAC_PARAMS)

if if_using_sac:
    # set up logger
    tmp_path = RESULTS_DIR + '/sac'
    new_logger_sac = configure(tmp_path, ["stdout", "csv", "tensorboard"])
    # Set new logger
    model_sac.set_logger(new_logger_sac)

```

```

{'batch_size': 128, 'buffer_size': 100000, 'learning_rate': 0.0001,
'learning_starts': 100, 'ent_coef': 'auto_0.1'}
Using cuda device
Logging to results/sac

```

```

[17]: trained_sac = agent.train_model(model=model_sac,
                                     tb_log_name='sac',
                                     total_timesteps=70000) if if_using_sac else None

```

```

day: 2892, episode: 110
begin_total_asset: 1000000.00
end_total_asset: 5060161.14
total_reward: 4060161.14
total_cost: 183863.98
total_trades: 58679
Sharpe: 0.848

```

```

=====
-----
| time/          |          |
|   episodes     | 4        |
|   fps          | 45       |
|   time_elapsed | 252      |
|   total_timesteps | 11572   |
| train/         |          |
|   actor_loss   | 325      |
|   critic_loss  | 189      |
|   ent_coef     | 0.108    |
|   ent_coef_loss | -105     |
|   learning_rate | 0.0001   |
|   n_updates    | 11471    |
|   reward       | 13.078064 |
-----
-----
| time/          |          |
|   episodes     | 8        |
|   fps          | 45       |
|   time_elapsed | 506      |

```

total_timesteps	23144
train/	
actor_loss	131
critic_loss	933
ent_coef	0.0342
ent_coef_loss	-153
learning_rate	0.0001
n_updates	23043
reward	18.951672

-----

day: 2892, episode: 120  
begin\_total\_asset: 1000000.00  
end\_total\_asset: 7178408.36  
total\_reward: 6178408.36  
total\_cost: 9118.39  
total\_trades: 54645  
Sharpe: 0.909

=====

time/	
episodes	12
fps	45
time_elapsed	759
total_timesteps	34716
train/	
actor_loss	65.3
critic_loss	30.1
ent_coef	0.011
ent_coef_loss	-169
learning_rate	0.0001
n_updates	34615
reward	-3.360691

-----

time/	
episodes	16
fps	45
time_elapsed	1013
total_timesteps	46288
train/	
actor_loss	42
critic_loss	37.3
ent_coef	0.00367
ent_coef_loss	-102
learning_rate	0.0001
n_updates	46187
reward	-0.06650945

-----

```

-----
| time/          |          |
|   episodes     | 20       |
|   fps          | 45       |
|   time_elapsed | 1266     |
|   total_timesteps | 57860    |
| train/         |          |
|   actor_loss   | 36.8     |
|   critic_loss  | 24.6     |
|   ent_coef     | 0.00209  |
|   ent_coef_loss | 3.27     |
|   learning_rate | 0.0001   |
|   n_updates    | 57759    |
|   reward       | 2.456344 |
-----

```

```

day: 2892, episode: 130
begin_total_asset: 1000000.00
end_total_asset: 6606023.19
total_reward: 5606023.19
total_cost: 2610.69
total_trades: 43440
Sharpe: 0.934
=====

```

```

-----
| time/          |          |
|   episodes     | 24       |
|   fps          | 45       |
|   time_elapsed | 1522     |
|   total_timesteps | 69432    |
| train/         |          |
|   actor_loss   | 26.4     |
|   critic_loss  | 7.65     |
|   ent_coef     | 0.00185  |
|   ent_coef_loss | -0.801   |
|   learning_rate | 0.0001   |
|   n_updates    | 69331    |
|   reward       | 2.1811397 |
-----

```

```
[18]: trained_sac.save(TRAINED_MODEL_DIR + "/agent_sac") if if_using_sac else None
```