

Commitment Paper: Reinforcement Learning for Stock Picking

Jeppe S. Johansen (pcv439)

May 2019

In finance CAPM is one of the pillars of modern stock picking and portfolio composition. CAPM is based on the idea that a single covariance matrix Ω and mean vector μ is the structural model generating individual returns of stocks. In this view all return are *i.i.d.*, and both μ and Ω can be reliably estimated due to classical results from finance. However it's a very strong assumption to assume, that something that obviously must a time series (the returns of a stock $r_t, r_{t+1}, r_{t+2}, \dots$), should not have any time dependency, and that μ and Ω , cannot change over time.

Simultaneously has statistical learning (commonly referred to as AI) had great progress in control problems in the form of reinforcement learning. Reinforcement learning has been famous for getting to super human levels in ATARI games, Chess, Go (Ancient board game, which is considerably harder than most other games from a computational view point), star craft, etc. In other words, reinforcement learning have shown itself capable of creating effective strategies in difficult environments.

In this paper i will try to assume a data generating process with the following properties:

$$\psi_t \sim \text{Bernoulli}(k) \tag{1}$$

$$\mu_t := \begin{cases} \mu_{t-1} & \psi_t = 0 \\ \mu_t \sim d_\mu & \psi_t = 1 \end{cases} \tag{2}$$

$$\Omega_t := \begin{cases} \Omega_{t-1} & \psi_t = 1 \\ \Omega \sim d_\Omega & \psi_t = 0 \end{cases} \tag{3}$$

$$r_t \sim N(\mu_t, \Omega_t) \tag{4}$$

More intuitively this can be understood as, at each time step i assume that the with some probability (denoted k in the equation) a structural break can occur. such a break will imply, that the a new μ and Ω will be drawn, and essence the returns r_t will now be drawn from a reparametrized distribution.

Assuming the DGP described above, i will set out to train an actor-critic network model, to do stock picking i.e. train a model, that can each time t picks a single stock to invest assets in, with the objective to maximize the expected long term value of the assets.

This will be done first by doing a simulation study with the proposed DGP, and the in the end taking the model to real data, looking at daily returns of 10 stocks for a period of +20 years.