Adaptive Expectations and Reinforcement Learning in Economics

Computational Economics

Dietmar Maringer

WWZ, University of Basel

Spring 2012

(Computational) Learning

Learning – what is it good for?

- process information
- react to environment
- improve performance
- discover regularities
- deal with new situations
- predict & anticipate

How?

- supervision
- examples
- conclusions
- experience
- "trial and error"
- feedback and reward

Supervised Learning

basic principle

- training data set $\leftrightsquigarrow$ validation data set
  - available: past cases of input and "correct" output
  - objective: train system to predict output
  - how: prediction error in training set to be minimized
  - check with validation data set to prevent over-fitting

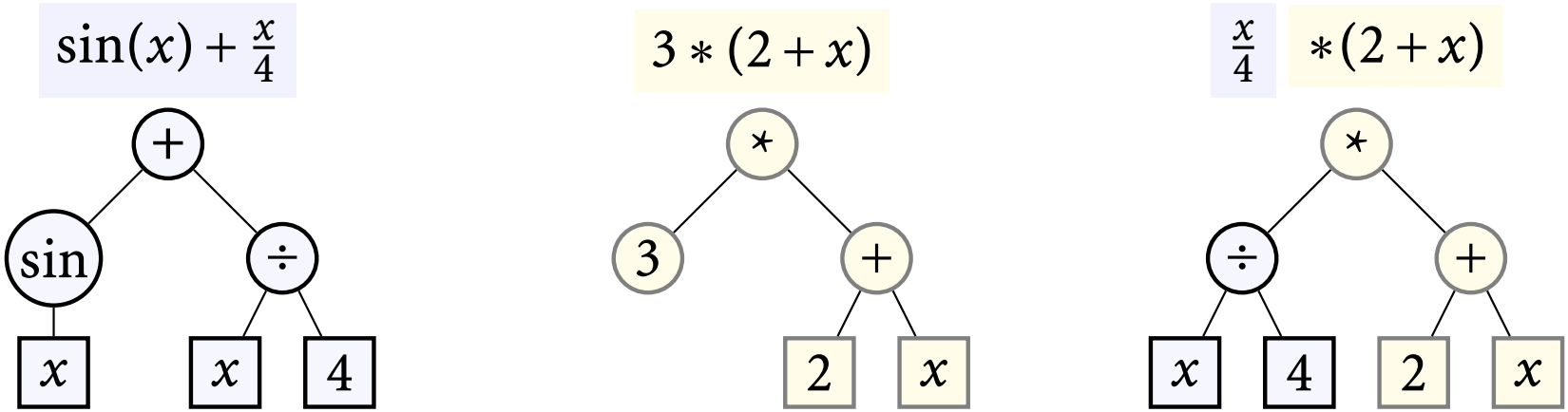
example: Artificial Neural Networks

- type of non-linear regression model
- applications in economics:
  - time series predictions
  - failure prediction / early warning systems
  - classification

Supervised Learning

Genetic Programming ([Koza, 1990] ☞*pdf*)

- lines of code / equations evolve over generations



Grammatical Evolution ([Ryan et al., 1998] ☞*pdf*)

- sequence of numbers, interpreted according to grammar ("lookup table" of functions and terminals)

Feedback and Reinforcement

Ingredients

- action – response
- feedback = response to agent's behavior
- feedback affects future behavior

Two lessons from psychology

- "power law of practice" (e.g., J.M. Blackburn, 1936)
  learning curves are initially step, but become flatter
- "law of effect" (E.L. Thorndike, 1898)
  choices with past good outcomes are likely to be repeated

$\Longrightarrow$ Reinforcement Learning (RL)

- successful actions are rewarded & encouraged
- unsuccessful actions are avoided

Reinforcement Learning in Economics

The main idea: [Erev and Roth, 1998] ☞pdf

- agent(s) i can choose between M different actions
- choosing action a results in receiving x_a
- ignorant of underlying principles, but observe action and response
- reinforcement function: $R(x_a) \geq 0$

Deciding and learning

- (initially equal) propensities: $q_{i,a}(t)$
- action and response observed, underlying principles unknown
- probability for choosing action k:

$$p_{i,a}(t) = \frac{q_{i,a}(t)}{\sum_j q_{i,j}(t)} \tag{1}$$

- updating of propensities after action k has been chosen:

$$q_{i,a}(t+1) = \begin{cases} q_{i,a}(t) + R(x) & a = k \\ q_{i,a}(t) & a \neq k \end{cases} \tag{2}$$

Reinforcement Learning in Economics

Example (RL 1)

- agent can choose between $M = 2$ actions
- rewards are $x_1 = 1$ and $x_2 = 2$
- reinforcement function: $R(x) = \exp(x)$

Example (RL 2)

- agent can choose between $M = 2$ actions
- rewards are $x_1 = 1 + \tilde{e}_1$ and $x_2 = 2 + \tilde{e}_2$ where $\tilde{e}_i \sim N(0, \sigma_e)$
- reinforcement function: $R(x) = \exp(x)$

Reinforcement Learning in Economics

Example (RL 3 (prisoners' dilemma))

- two villains, each committed one severe and one minor offence
- only the minor one was witnessed
- villains are interrogated separately
- their individual fines depend on whether they co-operate with the police:
  - if neither defeats: each is sentenced to 1 year of prison for the minor offence
  - if both defeat: each one is sentenced to 5 years of prison
  - if only one defeats: the co-operator is free, the other one gets 20 years

Generalized updating functions

$$q_{i,a}(t+1) = (1-\phi)\cdot q_{i,a}(t) + E_k(a, R(x)) \tag{3}$$

where $\phi \in \{0,1\}$ is a decay factor and $\varepsilon \in \{0,1\}$ is an experimentation / generalization factor in the experience function

$$E_k(a, R(x)) = \begin{cases} R(x_a)\cdot(1-\varepsilon) & a = k \\ R(x_a)\cdot\varepsilon/2 & a = k \pm 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$E_k(a, R(x)) = \begin{cases} R(x_a)\cdot(1-\varepsilon) & a = k \\ R(x_a)\cdot\varepsilon/(M-1) & a \neq k \end{cases} \tag{5}$$

Reinforcement Learning in Economics

Example (RL 4)

- consider a simple market with one supplier and one customer
- quantities for supply, q_S, and demand, q_D, depend on price p:

$$\text{supply function: } q_S = -1 + 2 \cdot p$$
$$\text{demand function: } q_D = 10 - p \quad (+\tilde{e})$$

- supplier sets price
- price must be multiple of £1

Reinforcement Learning in Finance

Learning in financial markets

- new information will influence prices
- magnitude of changes depends on
  - anticipated v completely unexpected
  - reliability
  - of temporary v permanent importance
- memory
- adapted expectations
- adapted behavior

Background: The Gordon Growth Model

Fundamental price of a dividend paying stock

$$S_0 = \frac{D_1 + \mathrm{E}(S_1)}{1+k} = \frac{D_1 + \frac{D_2 + \mathrm{E}(S_2)}{1+k}}{1+k} = \frac{D_1 + \frac{D_2 + \frac{D_3 + \mathrm{E}(S_3)}{1+k}}{1+k}}{1+k} = \cdots$$
$$= \sum_{t=1}^{T} \frac{D_t}{(1+k)^t} + \frac{\mathrm{E}(S_T)}{(1+k)^T}$$

for constant dividends

$D_t = D_{t-1} = D \quad \forall t$

$$S_0 = \sum_{t=1}^{\infty} \frac{D}{(1+k)^t} = \frac{D}{k}$$

for geometrically growing dividends

$D_t = D_{t-1} \cdot (1+g) \quad \forall t, 0 \leq g < k$

$$S_0 = \sum_{t=1}^{\infty} \frac{D_0 \cdot (1+g)^t}{(1+k)^t} = \frac{D_0(1+g)}{k-g} = \frac{D_1}{k-g}$$

Background: Rational Bubbles

- underlying idea: price differs from fundamental price S_t^f

$$S_t = \sum_{i=1}^{\infty} \frac{\mathrm{E}_t(D_{t+i})}{(1+k)^i} + B_t = S_t^f + B_t \tag{6}$$

- if agents are rational, market is perfect:

$$\begin{aligned} S_t &= \frac{\mathrm{E}_t(D_{t+1}) + \mathrm{E}_t(S_{t+1})}{1+k} \\ &= \frac{\mathrm{E}_t(D_{t+1}) + \sum_{i=2}^{\infty} \frac{\mathrm{E}_t(D_{t+i})}{(1+k)^i} + \mathrm{E}_t(B_{t+1})}{1+k} = \cdots \end{aligned} \tag{7}$$

- substituting (7) into (6): $\mathrm{E}_t(B_{t+m}) = B_t(1+k)^m$
- if bubble bursts at $t = \tau$: $B_\tau = 0$

Reinforcement Learning in Finance

Study and model by [Barsky and De Long, 1993] ☞*pdf*

- under the Gordon Growth Model:

$$S_0 = \frac{D}{k-g} \iff \underbrace{\ln(S_0)}_{\equiv s} = \underbrace{\ln(D)}_{\equiv d} - \ln(k-g) \tag{8}$$

- when D increases by 1%, then S_0 should also increase by 1%
  $\Delta d = .01 \Longrightarrow \Delta s = .01$
- Empirical study for the S&P 500, 1880–1988
  - results for 20 years horizon: $\Delta d = 1\% \Longrightarrow \Delta s = 1.6\%$
  - for 1949–1969: $\Delta d = 0.72\%$ while $\Delta s = 1.63\%$
  - volatility is 67% higher than volatility of "fundamental values" (i.e., values according to Gordon Growth Model)
- explanation: "adaptive expectations"

Adaptive expectations for dividend growth rates

- growth rates can change, i.e., g_t instead of g
- agents have expectations about the growth rates:

$$E_t(\Delta d_{t+1}) = g_t \tag{9}$$

- expectations are based on past expectations
- expectations are adapted according to forecasting error ε_t

$$E_t(\Delta d_{t+1}) = E_{t-1}(\Delta d_t) + (1-\theta)\varepsilon_t \tag{10}$$

$$\varepsilon_t = \Delta d_t - \underbrace{E_{t-1}(\Delta d_t)}_{=g_{t-1}} \tag{11}$$

then, after substituting and rearranging

$$E_t(\Delta d_{t+1}) = g_t = \theta g_{t-1} + (1-\theta)\Delta d_t \tag{12}$$

$$= g_0 + \sum_{i=0}^{t-1}(1-\theta)\varepsilon_{t-i} \tag{13}$$

Reinforcement Learning in Finance

Consequences for the price processes under adaptive expectations

- "fundamental values"

$$V_t = \frac{D_t}{k - g_t}$$

  are more volatile than under constant g
- long swings in prices can (partly) be explained

Example (Simulation)

Reinforcement Learning in Finance

"How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Processes" ([Timmermann, 1993] ☞*pdf*)

- dividend process:

$$\Delta \ln(D_t) = \mu + \varepsilon_t \quad \text{where } \varepsilon \sim \text{niid}(0, \sigma_\varepsilon^2), \mu + \sigma^2/2 < \ln(1+r)$$
$$\ln(D_t) = \ln(D_{t-1}) + \mu + \varepsilon_t$$
$$E(D_{t+\tau}) = D_{t-1} \exp\left((\mu + \sigma^2/2) \cdot \tau\right)$$

- under rational expectations:

$$S_t = \frac{\exp\left(\mu + \sigma^2/2\right)}{k - (\exp\left(\mu + \sigma^2/2\right) - 1)} D_t$$

Reinforcement Learning in Finance

Adaptive learning of parameters in Timmermann (1993)

- process is known, but not μ and σ
- estimated from previous n observations
  simplified:

$$\hat{\mu}_t = \frac{n-1}{n}\hat{\mu}_{t-1} + \frac{1}{n}\Delta\ln(D_t)$$

$$\hat{\sigma}_t^2 = \frac{n-1}{n}\hat{\sigma}_{t-1}^2 + \frac{1}{n}\left(\frac{n-1}{n}(\hat{\mu}_{t-1} - \Delta\ln(D_t))^2\right)$$

Example (Simulation)

Reinforcement Learning in Finance

Extension: [Timmermann, 1996] ☞*pdf*

- based on empirical evidence:
  dividends (in logs or in level) are trend stationary

$$\ln(D_t) = \rho \ln(D_{t-1}) + \gamma t + \mu + \varepsilon_t$$

where

$$\varepsilon \sim \text{niid}(0, \sigma_\varepsilon^2), \rho < 1 + r$$

- agents are no longer assumed to have access to the true model
- parameters can be estimated with recursive OLS
- "[...] agents' learning may generate predictability in stock returns and significantly increase the volatility of stock returns."
- see also [Pesaran and Timmermann, 1995] ☞*pdf* on the robustness and economic significance of the predictability of stock returns

Reinforcement Learning in Finance

. . . and the limits: [Lewellen and Shanken, 2002] ☞*pdf*

- extension to equilibrium model
- true parameters of the underlying model are unknown and have to be estimated
  $\Longrightarrow$ consider parameter uncertainty and consequences for prediction
- findings:
  - learning affects asset prices
  - in sample: re-identify predictability
  - out of sample: investors can neither perceive nor exploit this predictability

Exogenous v endogenous processes

- dividend processes are not entirely exogenous
- feedback mechanism from the prices
- self-referential learning
- [Marcet and Sargent, 1989] ☞*pdf*: $z_t = T(\rho_t)z_{t-1} + V(\rho_t)u_t$
  where z_t ...observed state variables used by the agents
  u_t ...shock
  ρ_t ...parameters (to be estimated recursively $\Longrightarrow \hat{\rho}_t$)
- application: VAR for dividends and prices:

$$\begin{aligned} \hat{D}_{t+1} &= \hat{\rho}_{1,t} D_t + \hat{\rho}_{2,t} P_t \\ \hat{P}_{t+1} &= \hat{\rho}_{3,t} D_t + \hat{\rho}_{4,t} P_t \end{aligned} \iff \begin{bmatrix} \hat{D}_{t+1} \\ \hat{P}_{t+1} \end{bmatrix} = \begin{bmatrix} \hat{\rho}_{1,t} & \hat{\rho}_{2,t} \\ \hat{\rho}_{3,t} & \hat{\rho}_{4,t} \end{bmatrix} \begin{bmatrix} D_t \\ P_t \end{bmatrix}$$

solution under assumption of constant parameters in:
Allan Timmermann, "Why do Dividend Yields Forecast Stock Returns?" *Economic Letters*, 46(2), 1994, 149–158.

Reinforcement Learning in Finance and Economics

The road ahead: interacting agents

- different types of market participants, e.g.,
  - smart money versus noise traders
  - fundamentalists versus chartists
  - trend followers versus contrarians
- agents learn from and mimic selected other agents
  [Becker, 1991] ☞*pdf*: *herding in restaurant choice*
- agents can change their own type
  [Kirman, 1993] ☞*pdf*: *artificial financial market*
- search and optimization strategies

Summary and critical assessment

- Why do we study artificial systems?
- What's the concept of reinforcement learning (RL)?
- How can RL be used to predict economic behavior?
- How can RL be used in financial modeling?
- What are the caveats of this approach?

Literature

- Barsky, R. B. and De Long, J. B. (1993).
  Why does the stock market fluctuate?
  The Quarterly Journal of Economics, 108(2):pp. 291–311.
- Becker, G. S. (1991).
  A note on restaurant pricing and other examples of social influences on price.
  Journal of Political Economy, 99(5):pp. 1109–1116.
- Erev, I. and Roth, A. E. (1998).
  Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria.
  The American Economic Review, 88(4):pp. 848–881.
- Kirman, A. (1993).
  Ants, rationality, and recruitment.
  The Quarterly Journal of Economics, 108(1):pp. 137–156.
- Koza, J. R. (1990).
  Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems.
  Technical Report STAN-CS-90-1314, Stanford University Computer Science Department.
- Lewellen, J. and Shanken, J. (2002).
  Learning, asset-pricing tests, and market efficiency.
  The Journal of Finance, 57(3):pp. 1113–1145.
- Marcet, A. and Sargent, T. J. (1989).
  Convergence of least-squares learning in environments with hidden state variables and private information.
  Journal of Political Economy, 97(6):pp. 1306–1322.
- Pesaran, M. H. and Timmermann, A. (1995).
  Predictability of stock returns: Robustness and economic significance.
  The Journal of Finance, 50(4):pp. 1201–1228.
- Ryan, C., Collins, J., and O'Neill, M. (1998).
  Grammatical evolution: Evolving programs for an arbitrary language.
  In *Lecture Notes in Computer Science 1391. First European Workshop on Genetic Programming 1998.*
- Timmermann, A. (1996).
  Excess volatility and predictability of stock prices in autoregressive dividend models with learning.
  The Review of Economic Studies, 63(4):pp. 523–557.
- Timmermann, A. G. (1993).
  How learning in financial markets generates excess volatility and predictability in stock prices.
  The Quarterly Journal of Economics, 108(4):pp. 1135–1145.