Investor Reinforcement Learning

Peiran Jiao *

Submitted to the ASFEE 6th Annual Meeting 2015

March 10, 2015

Extended Abstract

Conventional portfolio theories require investors to form subjective beliefs about probability distributions of future states. However, this can be so demanding in the real world that investors may instead resort to heuristic rules. This study focuses on one particular deviation from the conventional theories, investigating whether and how boundedly rational investors overweight experience when making decisions under uncertainty.

How experience, rewards or punishments, shapes subsequent behavior was first studied under reinforcement learning by Thorndike (1898), and later developed into models in psychology and economics,[1] with most applications in game theory explaining players' repeating the actions that are most rewarding in the past, even when the environment has changed.[2] The finance literature has documented that *more* experience induces better performance (Nicolosi, Peng and Zhu, 2009) and less disposition effect (Dhar and Zhu, 2006), and that *more rewarding* experience with IPO auctions (Kaustia and Knüpfer, 2008), 401(k) portfolios (Choi et al., 2009) and common stocks (Strahilevitz, Odean and Barber, 2011) increases an investor's subsequent demand for them, hurting their performance in general. However, little is known about the mechanism how investors learn from experience.

This study uses repeated investment tasks with feedback, where decisions using

*Postdoctoral Research Fellow, Department of Economics and Nuffield College, University of Oxford, Mannor Road Building, Mannor Road, OX1 3UQ, UK; Email: peiran.jiao@economics.ox.ac.uk; Office: +44 01865 278993; Cell: +44 07539 916593.

[1]See e.g. Suppes and Atkinson (1960); Selten and Stoecker (1986); Gilboa and Schmeidler (1995).

[2]See e.g. Erev and Roth (1998); Camerer and Ho (1999).

beliefs and reinforcements are both plausible, and estimates the weight participants place on each decision rule. The behavioral implications and predictive power of relevant models both with (Camerer and Ho, 1999; Nevo and Erev, 2012) and without (Kahneman and Tversky, 1979) learning will be evaluated. The reasons for choosing a lab experiment over other empirical methods are threefold: (1) an important element in the model, investors' beliefs, cannot be directly observed in the field; (2) when studying real-world prices, we lack a benchmark to conveniently disentangle the information value and reinforcement value in historical prices; (3) reinforcements and information can be directly manipulated in the lab for a strong test of the learning models.

The experiment has two stages. The first stage is the same across all conditions. It elicits attitudes towards risk and loss using the multiple-price-list approach adapted from Holt and Laury (2002). In the second stage, participants observe 4 hypothetical assets, whose prices are generated independently each period in a similar fashion as Weber and Camerer (1998). Every period, the direction of price change is determined by 4 equally-likely underlying processes, with the probability of price increase each period being 65%, 55%, 45% and 35% respectively. Price cannot stay unchanged. Then the price change magnitude is randomly drawn from $\{1, 3, 5\}$. Participants first observe 6 periods of price history. Then for the 20 subsequent periods they can choose one share of an asset to purchase each period (the buy task), which is automatically sold when the next period price is revealed. Another task is to predict the probability of price increase for each asset in each period (the predict task). Rewards in the buy task are calculated according to actual prices. Belief elicitations are incentivized using the quadratic scoring rule, corrected for risk attitudes (Offerman et al., 2009).

There are three conditions (A, B, and C), each containing three rounds (1, 2, and 3). Conditions A1, B1 and C1 use the same price sequences;[3] in B1, participants only have the predict task; in C1, they are endowed with experimental cash (EC) and only have the buy task; in A1 they do both tasks. A2 and B2 use a different set of price sequences,[4] and different initial endowments: EC plus an asset portfolio (A2) or additional EC (B2). In A2 the endowed assets are automatically sold after participants observe the 6-period price history, with any gain or loss added to their accounts. A3 and B3 use price sequences with the same ups and downs as A2 and B2, but different random draws of price change magnitudes from $\{1, 3, 5\}$. C2 and C3 respectively use the price sequences of A2 and A3, but participants have less

[3]Price Sequences Set 1 in Table 1.

[4]Price Sequences Set 2 in Table 1

Table 1: A Summary of the Experimental Conditions

	Round 1	Round 2	Round 3
Condition A *Endowment* *Price Seq.*	(A1) Buy+Predict 500 EC Price Seq. Set 1	(A2) Buy+Predict 500 EC+Assets Price Seq. Set 2	(A3) Buy+Predict Equivalent EC as A2 Price Seq. Set 2 with diff magnitudes
Condition B *Endowment* *Price Seq.*	(B1) Predict Only 500 EC Price Seq. Set 1	(B2) Buy+Predict Equivalent EC as A2 Price Seq. Set 2	(B3) Buy+Predict Equivalent EC as A2 Price Seq. Set 2 with diff magnitudes
Condition C *Endowment* *Price Seq.*	(C1) Buy Only 500 EC Price Seq. Set 1	(C2) Buy+Predict Equivalent EC as A2 Price Seq. Set 2 with less info	(C3) Buy+Predict Equivalent EC as A2 Price Seq. Set 2 with diff magnitudes and less info

information, in that they only know the four price-generating processes differ in the probability of price increase, but not the specific probabilities. The design is summarized in Table 1.[5] Three things are manipulated across conditions: initial endowment, size of price changes (reinforcements) and information.

The advantage of using price sequences predetermined in this manner is a clear benchmark for Bayesian beliefs and for the information value of historical prices. A Bayesian agent should care only about the number of ups, believe the sequence with more ups to be more likely to continue going up, and buy such shares in all periods, even in the low information condition. A quasi-Bayesian decision maker may additionally be influenced by the order of price changes,[6] but still not by the magnitudes. Choices that deviate from decision rules that merely rely on beliefs (Bayesian or quasi-Bayesian) can be easily detected.

The behavioral implication of simple reinforcement learning is the reluctance to shift away from an asset that brought gains and the excessive desire to avoid those that brought losses, which can be tested within each condition; this tendency should be weaker among those who observe but are not invested in the same price sequences. Controlling for beliefs, experienced outcomes should have no explanatory power for choices, unless individuals overweight experience.

[5]Note that the order of rounds will be randomized across participants.

[6]See e.g. Rabin (2002).

Based on the above arguments, the following hypotheses can be generated. Comparing A2 with B2, reinforcement learning predicts more choices by A2 participants of the assets that gained during the first 6 periods in their initial endowments. A comparison of Round 2 with Round 3 in each condition and across conditions can reveal between- and within-subject differences in the responses to price change magnitudes. The size of price changes should not affect a Bayesian agent at all but may affect subsequent choices made by reinforcement learners. A comparison of A1 *vis-à-vis* B1 and C1 can reveal whether and by how much reinforcements bias beliefs, and whether elicitation of beliefs influence buying decisions. A comparison between B2, B3 and C2, C3 will demonstrate the effect of information on learning. Nevo and Erev (2012) suggest that under incomplete information about the environment, decision makers may exhibit some distinct learning patterns.[7]

The belief and choice data will be used to structurally estimate model parameters using the Maximum Likelihood method. Specifically, the candidate models for this situation include those without learning, such as the expected utility theory and prospect theory (Kahneman and Tversky, 1979), and those with learning, such as the Experience-Weighted Attraction model (Camerer and Ho, 1999) and the I-SAW model (Nevo and Erev, 2012).[8]

If investors do learn from and overweight experience, this readily accommodates many empirical findings, such as the asymmetric effect of experience on the disposition effect in the domains of gains and losses (Feng and Seasholes, 2005), style investing (Barberis and Shleifer, 2003), category learning (Peng and Xiong, 2006), and the cohort effect (Malmendier and Nagel, 2009). A better understanding of individual investors' decision process can improve predictions of their behavior and market dynamics, inform the design of more efficient investor education, help brokerage firms improve their clients' performances, and increase market efficiency.

References

Barberis, Nicholas, and Andrei Shleifer. 2003. "Style investing." *Journal of Financial Economics*, 68(2): 161–199.

[7]For example, positive and negative surprises may both trigger changes.

[8]The EWA model, which is intended for interactions in games, will be adapted to this individual decision context.

Camerer, Colin, and Teck Ho. 1999. "Experience-weighted attraction learning in normal form games." *Econometrica*, 67(4): 827–874.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2009. "Reinforcement learning and savings behavior." *The Journal of finance*, 64(6): 2515–2534.

Dhar, Ravi, and Ning Zhu. 2006. "Up close and personal: Investor sophistication and the disposition effect." *Management Science*, 52(5): 726–740.

Erev, Ido, and Alvin E Roth. 1998. "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria." *American Economic Review*, 848–881.

Feng, Lei, and Mark S Seasholes. 2005. "Do investor sophistication and trading experience eliminate behavioral biases in financial markets?" *Review of Finance*, 9(3): 305–351.

Gilboa, Itzhak, and David Schmeidler. 1995. "Case-based decision theory." *The Quarterly Journal of Economics*, 605–639.

Holt, Charles A, and Susan K Laury. 2002. "Risk aversion and incentive effects." *American Economic Review*, 92(5): 1644–1655.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society*, 263–291.

Kaustia, Markku, and Samuli Knüpfer. 2008. "Do investors overweight personal experience? Evidence from IPO subscriptions." *The Journal of Finance*, 63(6): 2679–2702.

Malmendier, Ulrike, and Stefan Nagel. 2009. "Depression babies: Do macroeconomic experiences affect risk-taking?" National Bureau of Economic Research.

Nevo, Iris, and Ido Erev. 2012. "On surprise, change, and the effect of recent outcomes." *Frontiers in psychology*, 3.

Nicolosi, Gina, Liang Peng, and Ning Zhu. 2009. "Do individual investors learn from their trading experience?" *Journal of Financial Markets*, 12(2): 317–336.

Offerman, Theo, Joep Sonnemans, Gijs Van de Kuilen, and Peter P Wakker. 2009. "A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes." *The Review of Economic Studies*, 76(4): 1461–1489.

Peng, Lin, and Wei Xiong. 2006. "Investor attention, overconfidence and category learning." *Journal of Financial Economics*, 80(3): 563–602.

Rabin, Matthew. 2002. "Inference by Believers in the Law of Small Numbers*." *The Quarterly journal of economics*, 117(3): 775–816.

Selten, Reinhard, and Rolf Stoecker. 1986. "End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach." *Journal of Economic Behavior and Organization*, 7(1): 47–70.

Strahilevitz, Michal Ann, Terrance Odean, and Brad M Barber. 2011. "Once burned, twice shy: How naïve learning, counterfactuals, and regret affect the repurchase of stocks previously sold." *Journal of Marketing Research*, 48(SPL): S102–S120.

Suppes, Patrick, and Richard C Atkinson. 1960. *Markov learning models for multiperson interactions.* Vol. 5, Stanford University Press.

Thorndike, Edward L. 1898. "Animal intelligence: An experimental study of the associative processes in animals." *Psychological Monographs: General and Applied*, 2(4): i–109.

Weber, Martin, and Colin F Camerer. 1998. "The disposition effect in securities trading: An experimental analysis." *Journal of Economic Behavior and Organization*, 33(2): 167–184.