Review of Finance (2016) 20: 2015-2037
Advance Access publication: 1 August 2016

The Effect of Naïve Reinforcement Learning in the Stock Market*

SUNGHOON PARK and SUNGZOON CHO

Seoul National University

Abstract. Some investors who are subjected to naïve reinforcement learning create a spread between a stock's fundamental value and its equilibrium price. Naïve learners are more likely to repurchase a stock previously sold for a gain than one sold for a loss. This causes predictable equilibrium prices. We propose a proxy for the effect of naïve learning and show the profitability of a long-short strategy based on our proxy.

JEL Classification: G41, C51

1. Introduction

Naïve reinforcement learning is a simple probable principle for learning behavior in decision problems. The investors who follow the naïve reinforcement heuristics, 'Naïve Learners', pay more attention to their experiences of actions and payoffs than other factors that are considered by rational investors. Naïve learners are pleased to repeat the actions that was successful and avoid to repeat the investment decision which was painful.

In recent years, a number of researchers have presented the evidence of naïve learners and the characteristic of their investment decisions. Based on the findings of these works, we propose a proxy to estimate the influence of naïve reinforcement learning on the future stock return. We build long/short portfolio using the distinction between the proxy values of assets and find the average monthly return is more than 1.5% over 20 years in US Stock market. Our empirical results are economically and statistically significant even after controlling various risk factors such as size, value, profitability, investment pattern, turnover ratio, short-term return, and long-term return.

The rest of the article is organized as follows. Section 2 provides a brief

* The authors would like to thank an anonymous referee and the editor, for providing many valuable comments that helped us significantly improve the article. We are also grateful for useful comments from the seminar and conference participants at Seoul National University, INFORMS, Korea BI Data Mining Conference, Korea Industrial Engineering Conference.

introduction to the chracteristic of naïve learner. Section 3 presents the methodology how to estimate the effect of naïve reinforcment learning. In Section 4, we describe the empirical procedures and results under various conditions. Section 5 presents the conclusions.

2. Characteristics of Naïve Learner

In this section, we find charateristics of naïve learner from related literatures and make assumptions based on them.

Naïve reinforcement learning can be found in various asset markets. Choi et al. (2009) provide evidence of naïve reinforcement learning in 401(k) savings decisions. Strahilevitz et al. (2011) find the behavior in the decision of repurchasing individual stocks. Huang (2012) studies the subsequent purchase of stocks in industry sector level. The behavior is also observed in IPO market by Kaustia and Knupfer (2008). Malmendier and Nagel (2011) shows that naïve reinforcement learning influences the investment decision not only in the stock market but also in the bond market. Naive learners are widespread in the financial market. According to above findings, we assume that there are two types of agents in the stock market, sophisticated investors who have rational expectations and naive learners who are subjected to naïve reinforcement learning[1].

Change of the decision of investment in the same asset means that the experience alters the risk preference of the investor or changes the expectation of the asset. Risk-taking behavior can be changed by negative experiences in early life (Malmendier and Nagel, 2011) and financial crises (Guiso et al., 2013). However, there is no concrete evidence of change of risk preference caused by regular realization of stocks. It is proper to use standard economic model which assume that individual risk preference is stable across time (Stigler and Becker, 1977) in our assumption. Therefore, we assume that the naïve reinforcement learning changes the expectation of the investor in previously realized stock and alter their tendency to repurchase it.

Additional experiences continue to change the expectation of the investor. Erev and Roth (1995, 1998) describes the reinforcement learning model which stands for the incremental learning of cumulative experience. Their model well explain how economic agent evolves their propensity in a broad range of economics experiments. According to the basic concepts of their model, we assume that naïve learners cumulate their realizations (or payoffs) and weigh more on recent realizations than previous payoffs. Malmendier and

[1] Obviously, there are more than two types of investors in the stock market. We simplify the assumption to concentrate our empirical analysis on the naïve reinforcement learning.

Nagel (2011) also mentioned that more recent experiences the stronger effects. Intuitively, one-month-ago realization of a stock is probably more impact than ten-years-ago realization of the stock.

Camerer and Ho (1999) suggest experience-weighted attraction learning that treats both actual payoffs and forgone payoffs. For the simplicity, we assume that only actual and directly experienced outcomes affect future decisions.

3. Methodology

In this section, we propose a proxy to estimate the influence of naïve reinforcement learning on the future stock return based on the findings of section 1.

First, we show how demand of naïve learners skew the equilibrium price path of a stock. We assume the followings;

· The supply of risky stock is fixed and normalized to one.
· Public information about the stock arrives just prior to period-t round of trading.
· The fundamental value of the stock at period t is the fully rational price which reflect available public information and follows a random walk as Equation. (1).

$$F_{t+1} = F_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2) \tag{1}$$

Our basic model is a stripped-down overlapping generation model with two-period-lived agents (De Long et al. 1990; Samuelson 1958). Agents invest all the money to maximize the expected utility of total wealth in the second period. The only decision of agents is to choose a portfolio in the first period.

The economy contains two assets, a risk-free bond and a stock. The risk-free bond pays a fixed rate interest, r_f, after one period. The risk-free bond is in perfectly elastic supply. The price of a risk-free bond is always fixed at one. On the other hand, the stock is not inelastic supply: its supply is unchangeable and normalized to one unit at any period. The price of the stock in period t is denoted P_t.

In the market, there are two types of agents, sophisticated investors (denoted s) who have rational expectations and naive learners (denoted n) who are subjected to naïve reinforcement learning. We assume that the proportion of naive learners is ω and that all agents of a given type are identical. Both types of agents choose their portfolios in the first period (t) to maximize perceived expected utility with their own anticipation of the stock price in the second period (t+1).

The representative sophisticated investor in period t estimates the distribution of the stock price with the fundamental value of the stock, and so maximizes expected utility given that distribution. Expected value of stock price of the sophisticated investor at period t+1 is

$$E(P_{t+1}) = E(F_{t+1}) = F_t \tag{2}$$

and the variance of that

$$Var(P_{t+1}) = \sigma_\varepsilon^2 \tag{3}$$

The representative naive learner in period t estimates the distribution of the stock with the misperception of naïve learner and the fundamental value of the stock. Same as the sophisticated agents, naive learners thus maximize their expectation of utility given based on their belief of that distribution. The expected value of stock price of the naïve learner at period t+1 is

$$E(P_{t+1}) = E(F_{t+1} + \theta L_t) = F_t + \theta L_t \tag{4}$$

and the variance of that is

$$Var(P_{t+1}) = \sigma_\varepsilon^2 \tag{5}$$

L_t denotes the misperception of naïve learner. The misperception is learned by the cumulative experiences of previous investments until the beginning of period t. θ is a positive constant that measures the influence of the misperception, L_t, on price.

When the naive learners experienced positive outcomes more than negative outcomes, L_t is a positive value and the expected price of naïve learner higher than that of sophisticated agents. It means that naïve learners are more bullish than the sophisticated agents. On the other hands, L_t is a negative value and the expected price of naïve learner lower than that of sophisticated agents. It means that naïve learners are more bearish than the sophisticated agents.

Each agent's utility is a constant absolute risk aversion function of wealth at period t+1 (Exponential Utility):

$$U(W) = \frac{1}{\gamma}(1 - e^{-\gamma W}), \qquad \gamma > 0 \tag{6}$$

where γ is the coefficient of absolute risk aversion (γ > 0). If wealth is normally distributed, maximizing the expected utility of wealth is equivalent to maximizing

$$E(W) - \frac{1}{2}\gamma Var(W) \tag{7}$$

The sophisticated investor chooses the amount λ_t^s of the stock held to maximize the expected utility. The wealth at period t+1 is

$$W_{t+1} = (1+r)(W_t - \lambda_t^s P_t) + \lambda_t^s P_{t+1} \tag{8}$$

Expected wealth of the sophisticated investor at period t+1 is

$$E(W_{t+1}) = (1+r)W_t + \lambda_t^s (F_t - (1+r)P_t) \tag{9}$$

and the variance of the sophisticated investor is

$$Var(W_{t+1}) = (\lambda_t^s)^2 \sigma_\varepsilon^2 \tag{10}$$

From Equation (7), (9), and (10), the optimal amount of investment in the stock is

$$\lambda_t^s = \frac{F_t - (1+r)P_t}{\gamma \sigma_\varepsilon^2} \tag{11}$$

The naive learner chooses the amount λ_t^n of the stock to maximize expected utility. The wealth at period t+1 is

$$W_{t+1} = (1+r)(W_t - \lambda_t^n P_t) + \lambda_t^n P_{t+1} \tag{12}$$

Expected wealth of the sophisticated investor at period t+1 is

$$E(W_{t+1}) = (1+r)W_t + \lambda_t^n (F_t + \theta M_t - (1+r)P_t) \tag{13}$$

and the variance of the sophisticated investor is

$$Var(W_{t+1}) = (\lambda_t^n)^2 \sigma_\varepsilon^2 \tag{14}$$

Similar to the derivation of Equation (12), the optimal amount of investment in the stock is

$$\lambda_t^n = \frac{F_t + \theta M_t - (1+r)P_t}{\gamma \sigma_\varepsilon^2} \tag{15}$$

Since the supply of the stock is unchangeable and normalized to one unit at any period, the demand of the sophisticated agents and the demand of the noise agents must sum to one in equilibrium. At the equilibrium of supply and demand, the equilibrium price is

$$P_t^* = \frac{1}{1+r}(F_t + \omega\theta M_t - \gamma \sigma_\varepsilon^2) \tag{16}$$

where the proportion of naive learners is ω.

Above equation expresses the equilibrium price of the stock at period t as a function of the fundamental value and the misperception of naïve learner. When there is no naive learner (ω = 0), the equilibrium price of the stock should be $\frac{F_t-\gamma\sigma_\varepsilon^2}{1+r}$. The equilibrium price is lower than the fundamental value because of the uncertainty of the fundamental value at t+1. The equilibrium price is higher than the fundamental value only when there exist naïve learners and the naive learners have sufficient positive experience of previous investments. The higher the proportion of the naive learners (ω), the equilibrium price will be further from the expected stock price of sophisticated agents.

Second, we describe how to estimate the misperception of naïve learners. Naïve learners weigh more the directly experienced outcomes than the outcomes that are merely observed even if experience logically does not predict future success. If investors sell the stock for a loss, they feel pain and regret the past decision of purchasing the stock. This negative experience deters investors from later repurchasing the stock that they sold for a loss. On the other hand, if investors sell the stock for a gain, they are delighted to the past decision of purchasing the stock. This positive experience encourages investors from later repurchasing the stock that they sell for a gain. As a result, the more positive (negative) realized profits investors experienced, the more gain (loss) investors expect in the next investment of the same stock.

To estimate the misperception of naïve learner, L_t, we assume the followings[2];
· Risk preference of each agent is stable across time.
· Naïve reinforcement learning changes the propensity of agents to repurchase the stock.
· Only direct experience affects the propensity of agents to repurchase the stock.
· Positive experiences increase the propensity of repurchase, and negative payoffs decrease it.
· In the formation of misperception, naïve learners cumulate their realizations of past investment and weigh more on recent realizations than previous payoffs.

Naïve learner cumulates the experience of past realizations and put more weight on recent realized profits than on more distant realizations. L_t is interpreted as

$$L_t = M_{t-1} + (1-\delta)L_{t-1} \tag{17}$$

where δ is the depression rate which stands for the recency weighting scheme, ω is the proportion of naive learners, M_t is recency weighted aggregated realizations of all the shares which are sold at time t. We estimate the

[2] The assumptions are according to the findings from related works mentioned in section 1.

realization of all naïve learners as ωM_t. L_t is the aggregated misperception of all naïve learners in the stock.

$$M_t = \sum_{k=1}^{\infty} (P_t - P_{t-k}) V_{t-k} \left[\prod_{i=1}^{k-1} (1 - O_{t-k+i}) \right] O_t \tag{18}$$

In Equation (16), P_t is the stock price at end of period t, V_t is the trading volume in period t and O_t is the turnover ratio[3]. The derivation of M_t is provided in Appendix.B.

Finally, we propose the proxy for the return predictability. We can predict the expected change in the stock's price from t to t+1with the equilibrium price in Equation. (14). The derivation is provided in Appendix C. The expected return of the stock is

$$E\left(\frac{P_{t+1} - P_t}{P_t}\right) = \alpha\omega \frac{M_t}{S_t} - \alpha\delta \frac{L_t}{S_t} \tag{19}$$

where α is $\frac{\omega\theta n}{1+r_f}$, n is the number of outstanding shares, and St is the capitalization. Since the constant terms, ω, θ, δ, r_f, and n are positive, α is positive. New experience, M_t, and last misperception, L_t, is represented with money value. The big-cap stocks obviously have larger magnitudes of M_t and L_t than those of the small-cap stocks. To control the size effect, we divide them by the capitalization of the stock. There are two terms in Equation. (19). However, we only focus on the second term with L_t. First reason is the arrival time of information. L_t can be known just at the beginning of period t. L_t has the influence on the investment decisions of naive learners in period t. In the contrast with Lt, Mt has no influence on naive learners in period t because Mt is decided at the end of the period t. The other reason is that M_t is perfectly correlated to the unrealized profit of shareholders. As a result, M_t is not only a measure for reinforcement learning but also a measure for shareholder's psychology, such as the disposition effect (Grinblatt and Han, 2005).

According to Equation (19), the expected return decreases (increases) as the misperception of naive learners increases (decreases). The price moves in the opposite direction of the misperception of naive learners because the depreciation (δ) of misperception. We use the cap-adjusted value of depressed misperception as the proxy of the return predictability by naïve reinforcement learning. We name this proxy 'Proxy of Naïve Reinforcement Learning (PNLR)'.

[3] Turnover ratio in period t is calculated by dividing the trading volume traded in period t by the number of outstanding shares.

$$DMI_t = \delta \frac{L_t}{S_t} \tag{20}$$

It is reasonable to build a long/short portfolio which offsets a long position in stocks with low PNLRs by a short position in stocks with high PNLRs. Because of the return predictability of PNLR, we find that the average monthly returns of that long/short portfolio are positive and statistically significant in our empirical results.

4. Empirical Results

4.1 Data Description

We test the relation between the influence of naïve learner and the cross-section of expected returns in US Stock Market. We obtain stock data for the NYSE, AMEX, and NASDAQ from the CRSP (The Center for Research in Security Prices). The data set include the close price, the trading volume, and the number of outstanding shares for every month between December 1970 and December 2010. Price, dividend, shares, and volume data are historically adjusted for split events to make data directly comparable at different times during the history of a security.[4]

Based on the CRSP classification of the stocks, we exclude ADRs, REITs, and units of beneficial interest and use only ordinary firms. Following Jegadeesh and Titman (2001), we exclude all stocks priced below $5 at the beginning of the holding period and all stocks with market capitalizations that would place them in the smallest NYSE decile. We exclude these stocks to ensure that the results are not driven primarily by small and illiquid stocks or by bid-ask bounce. At the beginning of each month t, we estimate PNLR of each stock with its past ten years.

Table 1 shows the statistics for the selected stocks. Each column describes the following; the average number of ordinary firms listed in the CRSP database, the average total capitalization of ordinary firms listed in the CRSP database, the average number of the stocks with valid PNLRs, the average total capitalization of the stocks with valid PNLRs, the percentage of the stocks which have valid PNLRs, and the percentage of the total market value which have valid PNLRs.

---

[4] Split events always include stock splits, stock dividends, and other distributions with price factors such as spin-offs, stock distributions, and rights. Shares and volumes are only adjusted using stock splits and stock dividends. Split events are applied on the ex-distribution date.

Table 1. We obtained stock data for the NYSE, AMEX, and NASDAQ from the CRSP (The Center for Research in Security Prices). The data included the closing price, the trading volume, and the number of outstanding shares between December 1971 and December 2010. Based on the CRSP classification of the stocks, we excluded ADRs, REITs, and units of beneficial interest and included only ordinary firms in our empirical test. We exclude all stocks priced below $5 at the beginning of the holding period and all stocks with market capitalizations that would place them in the smallest NYSE decile. At the beginning of each month t, we select the stocks with their data of past ten years to estimate valid INLs (the influence of naïve learner). # Stocks and MV indicate the average number of ordinary firm stocks and the average total market value in the given period, respectively. % Stocks is the percentage of stocks about which have valid INLs, and % Cap is the percentage of the total market value have valid INLs.

Period	All Stocks		Selected Stocks		Proportion	
	# Stocks	MV ($1M)	# Stocks	MV ($1M)	% Stocks	% MV
1991-2000	6,166	7,415,637	1,777	5,966,837	28.8%	80.5%
2001-2010	4,603	13,067,758	1,944	10,871,313	42.4%	83.2%
1981-2010	5,385	10,241,698	1,860	8,419,075	34.5%	82.2%

According to Table 1, on average, 30.6% of stocks have valid PNLRs in the overall period. Since we use past ten year of prior data to estimate PNLR of each stock, about 70% of stocks does not participate in our empirical test. However, the stocks with valid PNLRs cover more than 81% of market capitalization of all ordinary firms in CRSP database. This means that the excluded stocks are very small stocks and does not contribute much to the stock market. We judge that the included stocks are sufficient to test our proposed methodology.

4.2 Empirical Test

The empirical tests mainly utilize the 1-month rolling strategy. At the beginning of each month t, we sort all stocks with their data of past ten years into deciles based on PNLR of each stock and create equally weighted portfolios from each decile. It is reasonable to build a long/short portfolio which offsets a long position in stocks with low PNLRs by a short position in stocks with high PNLRs. Because of the return predictability of PNLR, we find that the average monthly returns of that long/short portfolio are positive and statistically significant in our empirical results. If the stock market is efficient, the time series of returns of the portfolios should not earn any abnormal returns.

Fama and French (1993) find that three factors[5], market excess returns, capitalization and book-to-market equity, adequately explain the cross-section of returns on US stocks for the period 1963-1990. The regression is

$$r(t) - r_f(t) = \alpha + \beta_1 \left(MKT(t) - r_f(t) \right) + \beta_2 SMB(t) + \beta_3 HML(t) + \beta_4 RMW(t) + \beta_5 CMA(t) \tag{21}$$

RET is a series of portfolio returns, RF is the return on the one-month Treasury bill, and MKT is the value-weight return on all stocks. SMB (Small Minus Big) mimics the underlying risk factors in monthly returns related to size, and HML (High Minus Low) mimics the underlying risk factors in monthly returns related to book-to-market equity. SREV represents the underlying risk factors in monthly returns related to short-term reversal (long the stocks with low returns on the last month and short the stocks with high returns on the last month), MOM represents the underlying risk factors in monthly returns related to intermediate-term momentum (long the stocks with high 2-12 month returns and short the stocks with low 2-12 month returns), LREV represents the underlying risk factors in monthly returns related to long-term reversal (long the stocks with low 13-60 month returns and short the stocks with high 13-60 month returns), and TO represents the underlying risk factors in monthly returns related to turnover ratio (long the stocks with high turnover ratio and short the stocks with low ratio in recent 12 months).

Table. 2. reports the coefficients from the Fama-MacBeth regressions of PNLR by regressing them, cross-sectionally, on the stock's past short-, intermediate-, and long-term returns, logarithm of capitalization, and turnover ratio. (Fama and MacBeth, 1973). We perform cross-sectional regression every month, and adjust the standard errors for heteroscedasticity and autocorrelation using the Newey-West adjustment (Newey and West 1987). Except short-term return, the coefficient of other regressors have statistically significance. PNLRs are positively influenced by intermediate- and long-term returns. This is obvious since the calculation of PNLR is based on estimated realization of past ten years. Intermediate- and long-term returns are often used for the measurements of momentum and long-term reversal respectively.

Positive coefficients of intermediate- and long-term returns mean that our proposed measure, PNLR, does not simply blessed by momentum and long-term reversal. Those five factors can lead misinterpretation of the influence

---

[5] The monthly values of the three factors of the Fama-French Model and the risk-free rate are from Ken French's website. (http://mba.tuck.dartmouth.edu/ pages/faculty/ken.french/ data library.html)

of PNLR. In the following sections, we carefully control those factors to demonstrate the influence of PNLR.

Table 2. reports the coefficients from the Fama-MacBeth regressions of PNLR by regressing them, cross-sectionally, on the stock's past short-, intermediate-, and long-term returns, logarithm of capitalization, and turnover ratio. (Fama and MacBeth, 1973). Ret(1M) is the stock return for the prior one month, Ret(2M-12M) is the return for the previous 2-12 months, Ret(13M-36M) is the return for the previous 13-36 months, and log(Cap.) is the natural logarithm of the market capitalization. Turnover(1M) is the turnover ratio in the last month. Cross-sectional regressions were performed every month, and the standard errors were adjusted for heteroscedasticity and autocorrelation using the Newey-West adjustment (Newey and West 1987). t-Statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold. The R2 is the average R2 from the Fama-Macbeth cross-sectional regressions.

Decreasing Rate (δ)	δ=0.0001	δ=0.001	δ=0.01	δ=0.1
R_{-1}	0.440 (0.36)	0.402 (0.34)	0.152 (0.20)	0.016 (0.17)
$R_{-12:-2}$	**1.135** **(2.61)**	**1.099** **(2.63)**	**0.840** **(2.80)**	**0.361** **(3.80)**
$R_{-60:-13}$	**0.797** **(3.39)**	**0.762** **(3.43)**	**0.515** **(3.70)**	**0.077** **(3.40)**
log(Capitalization)	**0.492** **(4.86)**	**0.469** **(4.84)**	**0.302** **(4.60)**	**0.048** **(3.41)**
Turnover	**-10.035** **(-3.06)**	**-9.645** **(-3.09)**	**-6.720** **(-3.35)**	**-1.258** **(-4.32)**
Constant	**-3.258** **(-5.04)**	**-3.087** **(-5.06)**	**-1.909** **(-5.05)**	**-0.274** **(-3.84)**
$\overline{R^2}$	0.0230	0.0235	0.0300	0.0719

We compute abnormal returns based on a time-series regression of the portfolio excess returns using the seven factors which includes three factors of Fama and French (1993) model, short-term reversal, momentum, long-term reversal, and turnover ratio. We also examine alphas of seven factor model for each stock portfolio. If the returns of a portfolio are well explained by the seven factors, alpha should not have statistical significance. Otherwise, we can argue that the returns of a portfolio have unexplained factors without these factors in Equation. (21).

Table 3 reports the monthly average returns and alphas of main test. Consistent with the implication of PNLR, Table 3 shows that the average excess returns and alpha of decile portfolios increase monotonically as one moves from the top to the bottom decile. The baseline rolling strategy that is long the top 10% PNLR stocks and short the bottom 10% generates statistically significant returns and alphas over 30 years in US Stock market.

The returns and alphas of portfolios are changed depending on decreasing rate (δ).

The returns and alpha of less than 0.1 of decreasing rates show similar values at each level of decile. 0.1 of decreasing rate seems a little bit large to estimate the PNLR. It means that PNLR is not much sensitive to the change of decreasing rate. Our purpose of empirical test does not focus on finding the best value of decreasing rate, but target on showing the predictability of proposed proxy, PNLR. In the following sections, we just use the PNLR with 0.01 of decreasing rate.

4.3 Robustness

To validate the robustness of our empirical findings, we split the data set in three periods and test conduct same empirical test on each period. The first, second, and last ten years of our sample period present different portraits of the stock market. From 1981 to 1990, the average monthly market return is 1.125%, and the average monthly T-bill rate is 0.686% which is higher than the other periods. In addition, liquidity is low, and trading costs is high. In the second period, from 1991 to 2000, the average monthly market return is 1.415% which is higher than the other periods, and the average monthly T-bill rate is 0.386%. S&P500 has skyrocketed from 353.4 to 1320.28, which means the index increases almost 4 times in this period. The last period, 2001 to 2010, is quietly different from two previous periods. The market returns and risk-free rate is close to zero, and the tick size was decimalized.

Table 4 shows the average returns and the alphas of the long/short portfolios in three sub-periods. The average returns and alphas are statistically significant and positive in all sub-periods. However, the return of long/short portfolio in first period is much lower than the latter two periods and is even lower than average T-bill rate. Because the transaction cost in first period is greater than that of period, the transaction cost could disturb the process of equilibrium price formation. In first period, High T-bill rate also can be the reason of low return. When the risk-free rate is high, the risky assets are not attractive to both sophisticated investors and naïve learners. The lack of investors in stock market causes illiquidity, and then the process of price equilibrium cannot be achieved.

Table 3. We divide the stocks to ten decile portfolios at the beginning of each month based on the ranked values for naïve learning indices for NYSE stocks. Because most NASDAQ stocks are small-cap stocks, the split point is based on NYSE stocks. We weight the stocks in each portfolio equally, and use a 1-month rolling strategy. We also build the long/short portfolio that takes long position in bottom decile portfolio and short position in top decile portfolio. This table reports the average returns in excess of the risk-free rates and alphas of seven factor model. We present the result at each decreasing rate, 0.0001, 0.001, 0.01, and 0.1. Excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and statistical significance at the 5% level is indicated in bold.

	Average Excess Return				Alpha			
Decreasing Rate(δ)	δ=0.0001	δ=0.001	δ=0.01	δ=0.1	δ=0.0001	δ=0.001	δ=0.01	δ=0.1
10-High	0.361% (1.06)	0.355% (1.05)	0.284% (0.82)	0.517% (1.37)	**-0.480%** **(-4.02)**	**-0.481%** **(-4.02)**	**-0.524%** **(-4.33)**	**-0.313%** **(-2.13)**
9	0.431% (1.52)	0.401% (1.41)	0.447% (1.59)	**0.681%** **(2.42)**	**-0.432%** **(-4.55)**	**-0.458%** **(-4.87)**	**-0.418%** **(-3.98)**	-0.195% (-1.78)
8	**0.611%** **(2.29)**	**0.652%** **(2.41)**	**0.651%** **(2.48)**	**0.821%** **(3.21)**	**-0.263%** **(-2.86)**	**-0.242%** **(-2.63)**	**-0.237%** **(-2.82)**	-0.092% (-1.00)
7	**0.785%** **(3.00)**	**0.785%** **(3.04)**	**0.816%** **(3.21)**	**0.875%** **(3.54)**	-0.116% (-1.34)	-0.095% (-1.08)	-0.054% (-0.61)	-0.007% (-0.08)
6	**0.883%** **(3.44)**	**0.857%** **(3.35)**	**0.877%** **(3.51)**	**0.818%** **(3.27)**	0.017% (0.21)	-0.002% (-0.03)	0.004% (0.05)	-0.082% (-0.98)
5	**1.034%** **(4.06)**	**1.038%** **(4.11)**	**0.938%** **(3.68)**	**0.961%** **(3.72)**	0.122% (1.60)	0.118% (1.59)	0.010% (0.13)	0.012% (0.15)
4	**1.141%** **(4.54)**	**1.152%** **(4.55)**	**1.188%** **(4.68)**	**1.083%** **(4.08)**	**0.235%** **(2.96)**	**0.233%** **(2.89)**	**0.281%** **(3.48)**	0.177% (1.83)
3	**1.205%** **(4.28)**	**1.226%** **(4.36)**	**1.215%** **(4.19)**	**1.218%** **(4.05)**	**0.230%** **(2.41)**	**0.265%** **(2.80)**	**0.214%** **(2.23)**	**0.290%** **(2.38)**
2	**1.253%** **(3.83)**	**1.248%** **(3.79)**	**1.334%** **(4.01)**	**1.251%** **(3.66)**	**0.216%** **(2.14)**	0.201% (1.91)	**0.306%** **(2.70)**	**0.201%** **(1.44)**
1-Low	**1.935%** **(4.93)**	**1.930%** **(4.89)**	**1.964%** **(4.67)**	**1.775%** **(3.75)**	**0.838%** **(5.50)**	**0.830%** **(5.38)**	**0.823%** **(4.53)**	**0.661%** **(2.57)**
L/S	**1.574%** **(9.07)**	**1.575%** **(8.86)**	**1.680%** **(7.40)**	**1.258%** **(3.62)**	**1.318%** **(8.52)**	**1.311%** **(8.31)**	**1.346%** **(6.57)**	**0.973%** **(2.82)**

Table 4. This table reports factor loadings of Eq. (21) and alphas for the top (high) and bottom (low) decile portfolio and long/short portfolio. The dependent variable is the monthly excess return of the Treasury bill rate from the rolling strategy. The explanatory variables are the monthly excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold.

The first fifteen years and last ten years of our sample period present different portraits of the stock market. From 1986 to 2000, the average market returns were above one percent and the risk-free rate was quite high. In addition, liquidity was low, and trading costs, including commissions, were high. In the second half of our sample period, 2001 to 2010, the market returns and risk-free rates were lower, and the tick size was decimalized. This table shows the average excess returns and the alphas of the long/short portfolios based on the synergies between the two psychological effects. We compose a long/short stock portfolio that holds stocks with the top 10% PNRL and shorts stocks with the bottom 10% PNRL. Excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold. We used 252 business days as the mean lifetime of the forgetting function.

Period	Average Excess Return			Alpha			Avg. Market Return	Avg. T-bill Rate
	Low	High	L/S	Low	High	L/S		
1991 -2000	**1.679%** **(4.19)**	0.307% (0.67)	**1.372%** **(5.99)**	**0.541%** **(2.79)**	**-0.895%** **(-4.84)**	**1.436%** **(6.55)**	1.415%	0.386%
2001 -2010	**2.248%** **(3.03)**	0.260% (0.49)	**1.988%** **(5.08)**	**1.392%** **(5.02)**	**-0.308%** **(-2.00)**	**1.700%** **(5.18)**	0.319%	0.180%
1991 -2010	**1.964%** **(4.67)**	0.284% (0.82)	**1.680%** **(7.40)**	**0.823%** **(4.53)**	**-0.524%** **(-4.33)**	**1.346%** **(6.57)**	0.867%	0.283%

4.4 Double Sort

We check our empirical findings in different groups of stocks. We divide the stocks to five groups at the beginning of each month based on the ranked values for each well known risk factor. The factors are market capitalizations, average monthly turnover ratios, short-term returns, intermediate-term returns, and long-term returns. We examine the alphas of the long/short portfolios that takes buy bottom 20% PNLR stocks and sell top 20% PNLR stocks in each group. Table. 5 shows that alphas of long/short portfolios are statistically significant and positive alpha in every group of all factors. ‘Proxy of Naïve Reinforcement Learning’ maintains its predictability even if we control for various factors.

In the first column of Table 5, alphas of long/short portfolios increase as the level of market capitalization moves from the top to the bottom quintile. Institutional investors, who are trained not to make decisions on psychological effects, concentrate in large-cap stocks, and the proportion of naive learners in large-cap stocks is less than that in small-cap stocks.

Consequently, alpha of long/short portfolio in large-cap group is lower than that in small-cap group. The second column of Table 5 describes that alphas of long/short portfolios decrease as the turnover ratio of one year moves from high to low. High turnover ratio means a lot of realizations, and a lot of realizations enhance the magnitude of PNLR.

Table 5.a Double Sort (Capitalization)

Stocks are sorted by their market capitalization and then divided into five portfolios labeled 1 (Low), 2, 3, 4, and 5 (High). We examined the average excess returns and the alphas of the long/short portfolios based on the synergies between the two psychological effects at each level of capitalization. We use long/short portfolios based on positive and negative synergies between the two psychological effects. We compose a long/short stock portfolio that holds stocks with the top 10% PNRL and shorts stocks with the bottom 10% PNRL. Excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold. The results are based on the proposed estimators with a mean lifetime of 252 business days.

		PNLR (Proxy of Naïve Reinforcement Learning)					
		1-Low	2	3	4	5-High	1-5
		Average Excess Returns					
Capitalization	5-Big	**0.738%**	**0.712%**	0.446%	0.377%	0.032%	**0.706%**
		(2.19)	**(2.61)**	(1.77)	(1.49)	(0.12)	**(3.32)**
	4	**1.017%**	**0.788%**	**0.769%**	0.493%	0.018%	**0.999%**
		(2.60)	**(2.71)**	**(2.77)**	(1.83)	(0.06)	**(3.99)**
	3	**1.134%**	**1.035%**	**0.743%**	**0.796%**	0.180%	**0.954%**
		(2.92)	**(3.71)**	**(2.49)**	**(2.68)**	(0.53)	**(3.96)**
	2	**1.473%**	**1.070%**	**1.012%**	**0.946%**	0.285%	**1.188%**
		(3.23)	**(3.23)**	**(3.47)**	**(3.08)**	(0.79)	**(4.65)**
	1-Small	**2.447%**	**1.685%**	**1.345%**	**1.058%**	**0.824%**	**1.623%**
		(6.09)	**(5.15)**	**(5.37)**	**(4.32)**	**(2.61)**	**(8.23)**
		Alphas					
Capitalization	5-Big	-0.110%	-0.061%	**-0.323%**	**-0.267%**	**-0.503%**	0.392%
		(-0.77)	(-0.62)	**(-3.37)**	**(-2.88)**	**(-4.10)**	(1.90)
	4	-0.013%	-0.139%	-0.142%	**-0.375%**	**-0.700%**	**0.687%**
		(-0.07)	(-1.23)	(-1.22)	**(-3.01)**	**(-4.24)**	**(2.80)**
	3	-0.009%	0.051%	**-0.296%**	-0.169%	**-0.688%**	**0.679%**
		(-0.05)	(0.43)	**(-2.48)**	(-1.24)	**(-4.51)**	**(2.82)**
	2	0.178%	-0.095%	-0.016%	-0.095%	**-0.712%**	**0.890%**
		(0.87)	(-0.72)	(-0.14)	(-0.76)	**(-5.36)**	**(3.60)**
	1-Small	**1.339%**	**0.673%**	**0.509%**	**0.253%**	**-0.056%**	**1.395%**
		(7.12)	**(4.89)**	**(5.00)**	**(2.61)**	**(-0.44)**	**(7.74)**

Table 5.b Double Sort (Turnover)

Stocks are sorted by their turnover and then divided into five portfolios labeled 1 (Low), 2, 3, 4, and 5 (High). We examined the average excess returns and the alphas of the long/short portfolios based on the synergies between the two psychological effects at each level of turnover ratio. We use long/short portfolios based on positive and negative synergies between the two psychological effects. We compose a long/short stock portfolio that holds stocks with the top 10% PNRL and shorts stocks with the bottom 10% PNRL. Excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold. The results are based on the proposed estimators with a mean lifetime of 252 business days.

		PNLR (Proxy of Naïve Reinforcement Learning)					
		1-Low	2	3	4	5-High	1-5
		Average Excess Returns					
Turnover Ratio	5-High	**1.912%**	**1.404%**	**1.581%**	**0.923%**	0.182%	**1.731%**
		(3.54)	**(3.15)**	**(4.06)**	**(2.26)**	(0.40)	**(5.22)**
	4	**1.560%**	**1.419%**	**1.052%**	**0.703%**	0.394%	**1.166%**
		(3.86)	**(4.17)**	**(3.48)**	**(2.33)**	(1.21)	**(5.22)**
	3	**1.618%**	**1.176%**	**0.822%**	**0.715%**	0.339%	**1.279%**
		(4.40)	**(3.99)**	**(2.91)**	**(2.62)**	(1.23)	**(6.35)**
	2	**1.576%**	**1.012%**	**0.706%**	0.479%	**0.555%**	**1.021%**
		(4.96)	**(3.67)**	**(2.80)**	(1.94)	**(2.18)**	**(6.17)**
	1-Low	**1.420%**	**0.946%**	**0.808%**	**0.597%**	**0.591%**	**0.829%**
		(5.24)	**(4.34)**	**(4.03)**	**(2.81)**	**(2.66)**	**(6.48)**
		Alphas					
Turnover Ratio	5-High	**0.799%**	0.315%	**0.637%**	0.011%	**-0.580%**	**1.379%**
		(3.03)	(1.74)	**(3.98)**	(0.07)	**(-3.26)**	**(4.35)**
	4	**0.415%**	**0.314%**	0.060%	**-0.319%**	**-0.581%**	**0.996%**
		(2.40)	**(2.25)**	(0.51)	**(-2.69)**	**(-4.07)**	**(4.79)**
	3	**0.333%**	0.152%	-0.191%	**-0.270%**	**-0.573%**	**0.906%**
		(2.15)	(1.34)	(-1.68)	**(-2.69)**	**(-4.68)**	**(5.15)**
	2	**0.481%**	0.007%	**-0.216%**	**-0.406%**	**-0.273%**	**0.754%**
		(3.35)	(0.07)	**(-2.03)**	**(-3.37)**	**(-2.41)**	**(5.03)**
	1-Low	**0.502%**	0.135%	0.079%	-0.170%	-0.187%	**0.690%**
		(4.09)	(1.41)	(0.85)	(-1.63)	(-1.80)	**(5.83)**

4.5 Rolling Periods

We, then, analyzed the profitability of rolling investment strategies. For this purpose, we used a rolling portfolio approach, following Jegadeesh and Titman (1993) and Fama and French (1993). The resulting overlapping returns can be interpreted as the returns of a trading strategy that in any given month t holds a series of portfolios selected in the current month as well as in the previous k months, where k is the holding period. All stocks are value weighted within a given portfolio, and the overlapping portfolios are rebalanced at the end of each month to maintain equal weights. This approach makes it possible to estimate how long the profitability of the investment strategy will last. The empirical tests mainly utilize the 1-month rolling strategy. In the section on robustness, we also examine one-, two-, three-, six-, and twelve-month rolling strategies.

The time series of returns of the rolling portfolios track the monthly performance of each trading strategy. If the stock market is efficient, the trading strategy should not earn any abnormal returns. We compute abnormal returns based on a time-series regression of the portfolio excess returns using the three factors of the Fama and French (1993) model. Fama and French (1993) find that three factors[6], market excess returns, capitalization and book-to-market equity, adequately explain the cross-section of returns on US stocks for the period 1963-1990.

Table. 6. Long Horizon We use the rolling strategy suggested by Jegadeesh and Titman (1993) to examine how long these emotional effects influence stock prices. In any given month t, the strategies determine a series of portfolios that are selected in the current month as well as in the previous h-1 months, where h is the holding period. We use long/short portfolios based on the proposed estimator 'PNRL'. We compose a long/short stock portfolio that holds stocks with the top 10% PNRL and shorts stocks with the bottom 10% PNRL. The returns of long/short strategies were calculated for a series of portfolios that were rebalanced monthly to maintain value weights. We consider holding periods of 1, 2, 3, 6 and 12 months. This table shows the average excess returns and alphas of the long/short portfolios based on the synergies between the two psychological effects. Excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold. The results are based on the proposed estimators with a mean lifetime of 252 business days.

	Excess Return			Alpha		
	Low	High	L/S	Low	High	L/S
+1	**1.964%**	0.284%	**1.680%**	**0.823%**	**-0.524%**	**1.346%**
	(4.67)	(0.82)	**(7.40)**	**(4.53)**	**(-4.33)**	**(6.57)**
+2	**1.978%**	0.298%	**1.679%**	**0.837%**	**-0.501%**	**1.339%**
	(4.80)	(0.87)	**(8.59)**	**(4.97)**	**(-3.96)**	**(7.79)**
+3	**1.975%**	0.386%	**1.589%**	**0.831%**	**-0.444%**	**1.274%**
	(4.88)	(1.13)	**(9.21)**	**(5.26)**	**(-3.46)**	**(8.62)**
+6	**1.928%**	0.539%	**1.389%**	**0.800%**	**-0.339%**	**1.140%**
	(4.91)	(1.61)	**(9.96)**	**(5.63)**	**(-2.95)**	**(9.52)**
+12	**1.896%**	0.653%	**1.243%**	**0.785%**	**-0.259%**	**1.044%**
	(4.95)	(1.94)	**(10.34)**	**(5.75)**	**(-2.34)**	**(9.32)**

We use long/short portfolios based on the proposed estimator 'PNRL'. We

[6] The monthly values of the three factors of the Fama-French Model and the risk-free rate are from Ken French's website. (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html)

compose a long/short stock portfolio that holds stocks with the top 10% PNRL and shorts stocks with the bottom 10% PNRL. The monthly returns of long/short portfolio were calculated with value weights. We consider holding periods of 1, 2, 3, 6 and 12 months.

Table 6 shows the average excess returns and alphas of the proposed long/short strategy. Excess returns and alphas are presented as monthly percentages, t-statistics are listed below the coefficient estimates, and 5% statistical significance is indicated in bold. The results are based on the proposed estimators with a mean lifetime of 252 business days. The average rates of returns and the alphas are statistically significant at any holding periods. If the rolling period is longer than 3 months, the effect is slightly deceased disappears.

5. Conclusion

For psychological and emotional reasons, human beings do not always make decisions rationally. The vagarious nature of human behavior has been studied in psychology, economics and even finance.

Several papers in the finance literature have proposed behavioral theories to account for asset pricing anomalies. To provide support for their models' assumptions about investor behavior, these papers draw heavily on the experimental psychology literature, in which evidence of cognitive biases is abundant.

On the one hand, behavioralists contend that this evidence has been important in prompting researchers to consider heterodox explanations of market anomalies. On the other hand, skeptics argue that there exists so much of such evidence that behavioralists can "psycho-mine" the experimental psychology literature to find support for the particular set of assumptions that allow their models to match otherwise anomalous data. Contributing to the skeptics' argument, many of the behavioral theories rely on biases that are quite different from each other and often produce opposite conclusions about investor behavior. Not surprisingly, strong demand has emerged for empirical work that identifies which of the biases, if any, influences investor decisions. Even stronger is the demand to determine whether these biases are merely a curious aspect of certain market participants' behavior or whether they have important consequences for prices. This paper supplies evidence about both of these issues.

Empirical tests of behavioral models face a number of challenges. First, the models cannot be easily tested with aggregate data. As noted by Campbell (2000), "[Behavioral models] cannot be tested using aggregate consumption or the market portfolio because rational utility-maximizing investors neither consume aggregate consumption (some is accounted for by nonstandard investors) nor hold the market portfolio (instead they shift in and out of the stock market)". As a result, testing behavioral models is quite difficult without detailed information on the trading behavior of market participants. Unfortunately, given the issues of confidentiality associated with such data, availability of such information is generally quite low. An additional difficulty is that an investor's horizon, while highly ambiguous in most empirical settings, represents a key dimension in behavioral models. For instance, when fund managers are averse to losses, it is not clear whether their aversion relates to returns at the monthly, quarterly, or annual horizons, or even whether they view losses on positions taken recently as equivalent to losses on positions entered into years ago. Finally, even if biases can be identified in investor behavior, to demonstrate that this is more than just instances of noise trading, empirical tests must be positioned to identify a link between biases in individual trader behavior and overall prices.

Naïve reinforcement learning is a simple probable principle for learning behavior in decision problems. The investors who follow the naïve reinforcement heuristics, 'Naïve Learners', pay more attention to their experiences of actions and payoffs than other factors that are considered by rational investors. Naïve learners are pleased to repeat the actions that was successful and avoid repeating the investment decision which was painful.

In recent years, several researchers have presented the evidence of naïve learners and the characteristic of their investment decisions. Based on the findings of these works, we propose a proxy to estimate the influence of naïve reinforcement learning on the future stock return. We build long/short portfolio using the distinction between the proxy values of assets and find the average monthly return is more than 1.5% over 20 years in US Stock market. Our empirical results are economically and statistically significant even after controlling various risk factors such as size, value, profitability, investment pattern, turnover ratio, short-term return, and long-term return.

References

Choi, J. J., Laibson, D., Madrian, B., and Metrick, A. (2009) Reinforcement learning and savings behavior, *The Journal of Finance* **64**(6), 2515–2534.

Strahilevitz, M. A., Odean, T. and Barber, B. M. (2011) Once Burned, Twice Shy: How Naïve Learning, Counterfactuals, and Regret Affect the Repurchase of Stocks Previously Sold, *Journal of Marketing Research* **48**, S102-S120.

Kaustia, M., and Knüpfer, S. (2008) Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions, *The Journal of Finance* **63**(6), 2679-2702.

Malmendier, U., and Nagel, S. (2011) Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?, *Quarterly Journal of Economic*, **126**(1), 373–416.

Osili, U. O. and Paulson, A. (2014) Crises and Confidence: Systemic Banking Crises and Depositor Behavior, *Journal of Financial Economics* **111**, 646-660.

Huang, X. (2012) Industry investment experience and stock selection, Working paper.

De, S., Gondhi, N. R., and Pochiraju, B. (2010) Does sign matter more than size? An investigation into the source of investor overconfidence. SSRN Working paper.

J. Bradford De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990) The *Journal of Political Economy* **98**(4), 703-738.

Jegadeesh, N. and S. Titman (2001), Profitability of Momentum Strategies: An Examination of Alternative Explanations, *Journal of Finance* **56**(2), 699-720.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2013. “Time Varying Risk Aversion.” Booth Working Paper No. 13–64.

Stigler, George J., and Gary S. Becker. 1977. “De Gustibus Non Est Disputandum.” *American Economic Review* **67**(2): 76–90.

Redelmeier, D.A., and Tversky, A. (1992) On the Framing of Multiple Prospects, *Psychological Scienc,* **3**(3), 191-193.

Shafir, E., and LeBoeuf, R.A. (2002) Rationality, *Annual Review of Psychology* **53**, 491–517.

Seru, A., Shumway, T., and Stoffman, N. (2010) Learning by trading, *Review of Financial Studies* **23**(2), 705–739.

Singal, V., and Xu, Z. (2011) Selling winners, holding losers: Effect on fund flows and survival of disposition-prone mutual funds, *Journal of Banking and Finance* **35**, 2704–2718.

Shefrin, H., and Statman, M. (1985) The disposition to sell winners too early and ride losers too long: theory and evidence, *The Journal of Finance* **40**(3), 777–790.

Statman, M., Thorley, S., and Vorkink, K. (2006) Investor Overconfidence and Trading Volume, *Review of Financial Studies* **19**(4), 1531-1565.

Sullivan, K., and Kida., T. (1995) The effect of multiple reference points and prior gains and losses on managers’ risky decision making, *Organization Behavior and Human Decision Processes* **64**(1), 76–83.

Törngren, G. and Montgomery, H. (2004) Worse Than Chance? Performance and Confidence Among Professionals and Laypeople in the Stock Market, *The Journal of Behavioral Finance*, **5**(3), 148–153.

Tversky, A., and Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases, *Science* **185**(4157), 1124–1131.

Tversky, A., and Kahneman, D. (1981) The framing of decisions and the psychology of choice, *Science* **211**(4481), 453–458.

Tversky, A., and Kahneman, D. (1986) Rational choice and the framing of decision, *The Journal of Business* **59**(4), S251–S278.

Tversky, A., and Kahneman, D. (1992) Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* **5**(4), 297-323.

Weber, M., and Zuchel, H. (2005) How do prior outcomes affect risk attitude? Comparing escalation of commitment and the house-money effect. *Decision Analysis* **2**(1), 30-43.

Weber, M., and Camerer, C. F. (1998) The disposition effect in securities trading: An experimental analysis, *Journal of Economic Behavior and Organization* **33**, 167-184.

Winer, R. S. (1986) A reference price model of brand choice for frequently purchased products, *Journal of Consumer Research*. **13**(2), 250–256.

Wu, Y., Krehbiel, T., and Brorsen, B. W. (2011) Impacts of Tick Size Reduction on Transaction Cost, *International Journal of Economics and Finance* **3**(6), 57-65.

Appendix

A.1. Derivation of Equation (2)
According to Equation. (1),

$$\mu_{P_{t+1}} = E(F_{t+1}) = E(F_t + \varepsilon_{t+1}) = E(F_t) + E(\varepsilon_{t+1}) = F_t + 0 = F_t$$

A.2. Derivation of Equation (3)
According to Equation. (1) and (2),

$$\sigma^2_{P_{t+1}} = E\left[\left(P_{t+1} - E(P_{t+1})\right)^2\right] = E[(F_{t+1} - F_t)^2] = E[(\varepsilon_{t+1})^2] = \sigma^2_\varepsilon$$

A.3. Derivation of Equation (4)
According to Equation. (1),

$$\begin{aligned}\mu_{P_{t+1}} &= E(F_{t+1} + \theta L_t) = E(F_t + \varepsilon_{t+1} + \theta L_t) = E(F_t) + E(\varepsilon_{t+1}) + E(\theta L_t)\\ &= F_t + 0 + \theta L_t = F_t + \theta L_t\end{aligned}$$

A.4. Derivation of Equation (5)
According to Equation. (1) and (4),

$$\begin{aligned}\sigma^2_{P_{t+1}} &= E\left[\left(P_{t+1} - \mathrm{E}(P_{t+1})\right)^2\right] = \mathrm{E}\left[\left((F_t + \varepsilon_{t+1} + \theta L_t) - (F_t + \theta L_t)\right)^2\right] = \mathrm{E}[(\varepsilon_{t+1})^2]\\ &= \sigma^2_\varepsilon\end{aligned}$$

A.5. Derivation of Equation (9)
According to Equation. (2) and (8),

$$\begin{aligned}\mu_{W_{t+1}} &= E\left[\left(1 + r_f\right)W_t + \lambda^s_t\left(P_{t+1} - \left(1 + r_f\right)P_t\right)\right]\\ &= \left(1 + r_f\right)W_t + \lambda^s_t\left(E(P_{t+1}) - \left(1 + r_f\right)P_t\right)\\ &= (1 + r_f)W_t + \lambda^s_t\left(F_t - (1 + r_f)P_t\right)\end{aligned}$$

A.6. Derivation of Equation (10)
According to Equation. (3),

$$\begin{aligned}\sigma^2_{W_{t+1}} &= \mathrm{E}\left[\left(W_{t+1} - E(W_{t+1})\right)^2\right]\\ &= E\left[\left(\left(\left(1 + r_f\right)W_t + \lambda^s_t\left(P_{t+1} - \left(1 + r_f\right)P_t\right)\right) - E\left(\left(1 + r_f\right)W_t + \lambda^s_t\left(P_t - \left(1 + r_f\right)P_t\right)\right)\right)^2\right]\\ &= E\left[\left(\left(\left(1 + r_f\right)W_t + \lambda^s_t\left(P_{t+1} - \left(1 + r_f\right)P_t\right)\right) - \left(\left(1 + r_f\right)W_t + \lambda^s_t\left(E(P_t) - \left(1 + r_f\right)P_t\right)\right)\right)^2\right]\\ &= \mathrm{E_t}\left[\left(\lambda^s_t P_{t+1} - \lambda^s_t E(P_t)\right)^2\right]\\ &= (\lambda^s_t)^2 \mathrm{E_t}[(P_{t+1} - E(P_t))^2]\\ &= (\lambda^s_t)^2 \sigma^2_{P_{t+1}} = (\lambda^s_t)^2 \sigma^2_\varepsilon\end{aligned}$$

A.7. Derivation of Equation (12)
The optimal solution of λ^s_t makes that the first derivative of Equation. (11) equals to zero.

$$\max_{\lambda^s_t}\left[\left(\left(1 + r_f\right)W_t + \lambda^s_t\left(F_t - \left(1 + r_f\right)P_t\right)\right) - \frac{1}{2}\gamma(\lambda^s_t)^2\sigma^2_\varepsilon\right]$$

$$\left(F_t - (1 + r_f)P_t\right) - \gamma\lambda^s_t\sigma^2_\varepsilon = 0$$

$$\lambda_t^s = \frac{F_t - (1 + r_f)P_t}{\gamma\sigma_\varepsilon^2}$$

A.8. Derivation of Equation (13)

According to Equation. (4), Expected value of final wealth of the naïve learner at period t+1 is

$$\begin{aligned}\mu_{W_{t+1}} &= E\left[\left(1 + r_f\right)W_t + \lambda_t^s\left(P_{t+1} - \left(1 + r_f\right)P_t\right)\right] \\ &= \left(1 + r_f\right)W_t + \lambda_t^s\left(E(P_{t+1}) - \left(1 + r_f\right)P_t\right) \\ &= (1 + r_f)W_t + \lambda_t^s\left(F_t + \theta L_t - (1 + r_f)P_t\right)\end{aligned}$$

According to Equation. (5), the variance of final wealth of the naïve learner at period t+1 is

$$\begin{aligned}\sigma_{W_{t+1}}^2 &= E\left[\left(W_{t+1} - E(W_{t+1})\right)^2\right] \\ &= E\left[\left(\left(\left(1 + r_f\right)W_t + \lambda_t^n\left(P_{t+1} - \left(1 + r_f\right)P_t\right)\right)\right.\right. \\ &\qquad \left.\left. - E\left(\left(1 + r_f\right)W_t + \lambda_t^n\left(P_t - \left(1 + r_f\right)P_t\right)\right)\right)^2\right] \\ &= E\left[\left(\left(\left(1 + r_f\right)W_t + \lambda_t^n\left(P_{t+1} - \left(1 + r_f\right)P_t\right)\right)\right.\right. \\ &\qquad \left.\left. - \left(\left(1 + r_f\right)W_t + \lambda_t^n\left(E(P_t) - \left(1 + r_f\right)P_t\right)\right)\right)^2\right] \\ &= \mathrm{E_t}\left[\left(\lambda_t^n P_{t+1} - \lambda_t^n E(P_t)\right)^2\right] \\ &= (\lambda_t^n)^2 \mathrm{E_t}[(P_{t+1} - E(P_t))^2] \\ &= (\lambda_t^n)^2 \sigma_{P_{t+1}}^2 = (\lambda_t^n)^2 \sigma_\varepsilon^2\end{aligned}$$

Maximizing the utility of the representative sophisticated agent equivalent to maximize

$$\mu - \frac{1}{2}\gamma\sigma^2 = \left((1 + r_f)W_t + \lambda_t^n\left(F_t + \theta L_t - (1 + r_f)P_t\right)\right) - \frac{1}{2}\gamma(\lambda_t^n)^2\sigma_\varepsilon^2$$

The optimal solution of λ_t^s makes that the first derivative of Equation. (11) equals to zero.

$$\left(F_t + \theta L_t - (1 + r_f)P_t\right) - \gamma\lambda_t^n\sigma_\varepsilon^2 = 0$$

$$\lambda_t^n = \frac{F_t + \theta L_t - (1 + r_f)P_t}{\gamma\sigma_\varepsilon^2}$$

A.9. Derivation of Equation (14)

Since the supply of the stock is unchangeable and normalized to one unit at any period, the demand of the sophisticated agents and the demand of the noise agents must sum to one in equilibrium. The proportion of naive learners is ω and the proportion of sophisticated investors is (1- ω).

$$(1 - \omega)\lambda_t^s + \omega\lambda_t^n = 1$$

From Equation. (12) and (13),

$$(1 - \omega)\frac{F_t - (1 + r_f)P_t^*}{\gamma\sigma_\varepsilon^2} + \omega\frac{F_t + \theta L_t - (1 + r_f)P_t^*}{\gamma\sigma_\varepsilon^2} = 1$$

$$\frac{F_t - (1 + r_f)P_t^* + \omega\theta L_t}{\gamma\sigma_\varepsilon^2} = 1$$

From the above equation, the equilibrium price is

$$P_t^* = \frac{1}{1 + r_f}(F_t + \omega\theta L_t - \gamma\sigma_\varepsilon^2)$$

B. Derivation of Equation (16)

The stock price at end of period t, P_t, can be used as a proxy for the price of all transactions that occurred in the period, even though the actual transactions occurred during round of trading in period t. There is no reason to expect the price, P_t, to bias the results one way or the other. Realized profit of the share which are purchased in period t-k and sold in period t is

$$P_t - P_{t-k}$$

We assume that the holding period of a share is at least one period even though there are day-traders in real world. By the assumption, the number of shares which are purchased in period t and hold to next period is same as the trading volume in period t, V_t. If the turnover ratio, O_t, is the probability of realization, The number of shares which are purchased at time t-1 and sold at time t is

$$V_{t-1}O_t$$

,and the number of shares which are purchased at time t-1 and hold at time t is

$$V_{t-1}(1 - O_t)$$

The number of shares which are purchased at time t-k and sold at time t is

$$V_{t-k}(1 - O_{t-k+1})(1 - O_{t-k+2}) \dots (1 - O_{t-2})(1 - O_{t-1})O_t$$

$$V_{t-k}\left[\prod_{i=1}^{k-1}(1 - O_{t-k+i})\right]O_t$$

With above equation, we can estimate the total realized profit of the shares which are purchased on time t-k and sold at time t as

$$(P_t - P_{t-k})V_{t-k}\left[\prod_{i=1}^{k-1}(1 - O_{t-k+i})\right]O_t$$

Finally, we cumulate all the realizations in period t as

$$M_t = \sum_{k=1}^{\infty}(P_t - P_{t-k})V_{t-k}\left[\prod_{i=1}^{k-1}(1 - O_{t-k+i})\right]O_t$$

C. Derivation of Equation (17)

$$E\left(\frac{P_{t+1} - P_t}{P_t}\right)$$
$$= \frac{1}{P_t}E\left[\left(\frac{1}{1 + r_f}(F_{t+1} + \omega\theta L_{t+1} - \gamma\sigma_\varepsilon^2)\right) - \left(\frac{1}{1 + r_f}(F_t + \omega\theta L_t - \gamma\sigma_\varepsilon^2)\right)\right]$$
$$= \frac{1}{1 + r_f}\frac{1}{P_t}[E(F_{t+1} - F_t) + \omega\theta E(L_{t+1} - L_t)]$$
$$= \frac{1}{1 + r_f}\frac{1}{P_t}\left[E(\varepsilon_{t+1}) + \omega\theta E\big((\omega M_t + (1 - \delta)L_t) - L_t\big)\right]$$
$$= \frac{1}{1 + r_f}\frac{1}{P_t}\omega\theta(\omega M_t - \delta L_t)$$
$$= \frac{1}{1 + r_f}\frac{n}{nP_t}\omega\theta(\omega M_t - \delta L_t)$$
$$= \frac{1}{1 + r_f}\frac{n}{S_t}\omega\theta(\omega M_t - \delta L_t)$$
$$= \frac{\omega^2\theta n}{1 + r_f}\omega\frac{M_t}{S_t} - \frac{\theta n}{1 + r_f}\frac{\delta}{\omega}\frac{L_t}{S_t}$$
$$= \alpha\omega\frac{M_t}{S_t} - \alpha\delta\frac{L_t}{S_t}$$

where α is $\frac{\omega\theta n}{1+r_f}$, n is the number of outstanding shares, and St is the capitalization.