Make hay while the sun shines: an empirical study of maximum price, regret and trading decisions*

Julia Brettschneider† Giovanni Burro‡ Vicky Henderson§

May 30, 2020

Abstract

We test how regret and maximum price shape the propensity to realise stocks, on a large discount brokerage dataset containing US households' trading records between 1991 and 1996. First, we reject the hypothesis that investors stop at a threshold, which means they do not sell a stock on the day when the maximum or the minimum unfold. Only 31.6% of the gains are sold on the day when the maximum realises and 25.8% of the losses are sold on the day when the minimum realises. We find that more sophisticated and younger investors are more likely to follow a threshold strategy. Second, we find that investors are more likely to sell a stock for a gain in a moment closer in time to maximum occurrence and at a price further from the running maximum price of the stock in the investment episode. This cannot be rationalised through dynamic Regret Theory when the agent is only focused on past forgone prices but can be rationalised through anticipated regret, with the agent anticipating future potential regret. The propensity to sell a gain steadily declines a short time after the maximum occurred. We suggest that traders might regret not selling at a time close to the maximum day and hold onto the stock if a long time has passed.

Keywords: Regret Theory, Trading, Threshold, Maximum Price

JEL Codes: C55, D90, G40

*We thank Terrance Odean for making the data available. We thank David Hobson, Dario Spanò, Jo Kennedy, Daniel Read, Michele Fioretti, David Firth, Frédéric Schneider, Alessandro Bucciol (discussant), Gerrit Antonides (discussant) and seminar participants at the University of Warwick (Bridges, YRM), University of Nottingham (CeDEx), Loughborough University (BEDS 2019), Rome (IISES 2018), University of Verona (Economics Department), University of Naples (CSEF), University of Rome Sapienza (YESS), Dublin (IAREP/SABE 2019), Dublin (2019 SABE ECRW), Bertinoro (WEEE 2019), Amsterdam (SPUDM 2019), Trento (JDMx 2019), Milan (ADA 2019), Naples (SEET 2020) for their comments and suggestions. Giovanni Burro gratefully acknowledges the support of the Leverhulme Trust, through the Bridges DTC.

†University of Warwick, Department of Statistics; julia.brettschneider@warwick.ac.uk

‡University of Warwick, Department of Statistics; g.burro@warwick.ac.uk

§University of Warwick, Department of Statistics; vicky.henderson@warwick.ac.uk

1 Introduction

Our work contributes to two broad fields: decision making and empirical analysis of the behaviour of financial investors. First, we test an extension of Regret Theory (RT), one of the most successful theories of decision under risk (Loomes and Sugden, 1982; Bleichrodt and Wakker, 2015; Strack and Viefers, 2019). Second, we add to the literature on the behavior of individual investors (see Barber and Odean, 2013 for a comprehensive review of the topic). We test an application of Regret Theory to dynamic decisions (Strack and Viefers, 2019), in the context of financial trading decisions. There, Strack and Viefers (2019) show that an Expected Utility (EU) agent would always stop at an optimal threshold. It means that she would never sell at a price lower than a price at which she previously had the possibility to sell the stock. A Regret Theory agent does not necessarily stops at a threshold when deciding to sell a stock and her propensity to sell increases as the price of the stock increases and decreases as the distance from the past maximum, since purchase, increases. We test those predictions on the LDB dataset (Odean, 1998, 1999; Dhar and Zhu, 2006; Barber and Odean, 2013).

First, we test if investors stop on the day when the maximum realises for gains and on the day when minimum realises for losses and we find that only 31.6% of trading episodes in the gain domain and 25.8% of trading episodes in the loss domain are stopped on the day when maximum and minimum unfold, respectively. The discrepancy in the two figures is a corollary of the the disposition effect, the higher propensity to realise gains with respect to losses (Shefrin and Statman, 1985; Odean, 1998).

Second, we characterize those traders who are more likely to stop on the day when maximum or minimum realise (stop at a threshold) and build some links with the literature on trading decisions and individual characteristics (Shapira and Venezia, 2001; Dhar and Zhu, 2006; Seasholes and Zhu, 2010; Korniotis and Kumar, 2011, 2013). We adopt a negative binomial regression to model the rate at which investors stop at a threshold and find that sophisticated investors and active traders are more likely to follow a threshold strategy and affluent and older investors are less likely to follow a threshold strategy. Males are more willing to realise losses at a threshold than females.

Most importantly, we investigate the impact of the past maximum of a stock, since the stock was purchased, on the propensity to sell that stock. We model the time to sell using a proportional hazard model (Cox, 1972). This method has the advantage of assessing the impact of the covariates over the entire time axis. A proportional hazard model incorporates all the information accumulated in time for a given episode. We take into account both the distance in price from the past maximum and the distance in time from the maximum realisation day and the return of the stock. Consistent with Strack and Viefers (2019) we find that investors are more likely to realise a gain, the higher is the return.

In a Dynamic Regret Theory setting, when the agent is only concerned with regret about past forgone decisions, the propensity to sell decreases as the distance from the past maximum increases. The idea is that, ceteris paribus, if the price of the stock is further from the past maximum, the regret is higher and the investor will be less likely to stop. Figure 1 helps clarifying the point. The Utility of an Expected Utility agent only depends on the level of the price, hence it would be the same at any time the price hits the level highlighted in red. A Regret Theory agent would experience a higher Utility by selling the stock the first time the price reaches the price in red (left arrow in blue) than the second time (right arrow in blue). However, we find that the relationship of the propensity to realise a gain with the distance in price from the past maximum follows an inverse U-shape. It peaks when the distance in price is low but it decreases when it gets very low.

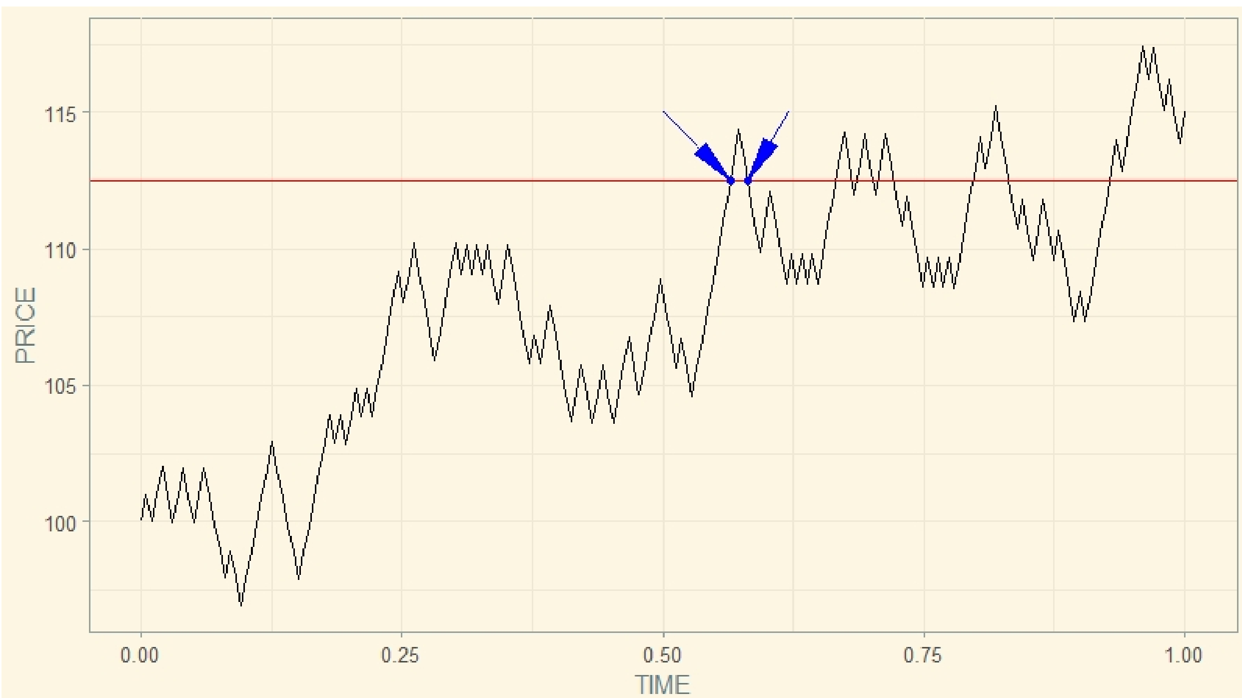

Figure 1: **Example of a threshold investment strategy.** Example of a threshold investment strategy. The upper threshold is highlighted in red. The Utility of an Expected Utility agent only depends on the level of the price. A Regret Theory agent would experience a higher Utility by selling the stock the first time the price reaches the threshold (left arrow in blue) than the second time (right arrow in blue).

We also take into account the distance in time from the past maximum. We find a strong and clear pattern: investors are less likely to sell a winning stock, the further in time the maximum price occurred. Moreover, we investigate the relation between distance in price and distance in time from the past maximum and find that, when the distance in time is low, investors are more likely to realise a gain, the higher is the distance in price from the past maximum. We suggest that a panic effect is the main force leading to the realisation of a stock. This finding can be rationalized with the experimental evidence in Fioretti et al. (2018). Investors anticipate regret about the future and decide to realise the stock to avoid experiencing an even higher regret in case the stock price keeps decreasing.

Regret theory has already been used to explain several phenomena in finance and economics: asset pricing and portfolio choice (Gollier and Salanié, 2006; Muermann et al., 2006, personal insurance market (Braun and Muermann, 2004), why people tend to invest too little in stocks (Barberis et al., 2006), hedging with respect to currency risk (Michenaud and Solnik, 2008) and the disposition effect (Muermann and Volkman, 2006).

We are the first to perform a test of dynamic regret on empirical data. There have been three attempts to study dynamic regret in a laboratory setting. Our closest predecessor is Strack and Viefers (2019). They extend Regret Theory to dynamic trading decisions and test their predictions in a laboratory experiment. As we said, they show that an Expected Utility agent would find it optimal stopping at a threshold and her propensity to sell would not be influenced by the distance from the past maximum. A Regret Theory agent does not necessarily stops at a threshold and her propensity to sell would be increasing in the price of the stock and decreasing in the price distance from the past maximum. Their predictions are respected by subjects in the laboratory. They find that agents do not follow a threshold strategy, which means they stop at a level where they decided to continue before. Agents are more willing to stop the higher the level of the price and they are less willing to stop the higher the maximum level of the price. Fioretti et al. (2018) perform a stock market experiment, where the participants know beforehand whether they will observe the future prices after they sell the asset or not. When future prices are available, investors avoid regret about expected after-sale high prices (future regret). Descamps et al. (2016) study how regret influences information sampling. Participants deviate from the optimal strategy in a systematic manner: information is either mostly over-sampled or mostly under-sampled, depending on the cost of information.

There are several papers linked to the idea that the maximum point of a price process has an effect on agents' behaviour. First of all, Grinblatt and Keloharju (2001) show that high recent past returns of stocks tend to increase the propensity to sell of the investors and the propensity to sell is higher when a stock hits its last month maximum price. Heath et al. (1999) show that the exercise of an option is higher when the price of the underlying stock is above its last year's peak. Barber and Odean (2008) and Huddart et al. (2009) find that trading volume is high around both last year maximum and minimum. Finally, an experimental paper by Baucells et al. (2011) investigates the impact of the highest level in a stream of payoffs on the formation of the reference point and find that the maximum price has a positive but weak effect on the formation of the reference point.

2 Data and methodology

The Large Discount Brokerage dataset (referred to as the *LDB Dataset*) is well established in the economic literature (Odean, 1999; Barber and Odean, 2000, 2001; Coval et al., 2005; Ivković and Weisbenner, 2005; Ivković et al., 2008; Seasholes and Zhu, 2010; Korniotis and Kumar, 2011, 2013; Huang, 2019). It contains information on trading activities of individual households in the USA from January 1991 to November 1996. We use two different samples of the data for the two sections of our analysis. We are going to explain the details of all our choices. Some of the data preparation steps are common between the two specifications and we will describe them here. We obtain price data from the CRSP (Centre for Research in Security Prices) of WRDS (Wharton Research Data Services). We exclude those stocks for which we were not able to recover price information. We remove investor-stocks records if at least one of the entries has negative commissions (which may indicate that the transaction was reversed by the broker). We remove from the sample investor-stocks that include short-sale transactions or that have positions that were opened before the starting point of our dataset. We remove those trades where the buy and sell dates coincide (Ivković et al., 2005; Ben-David and Hirshleifer, 2012).

The starting point of an investment is the first time an investor buys a stock or any time she buys it without the stock being present in the bank account at that time. The end point of an investment is the first sale date after that buy date as the end point of the investment (Shapira and Venezia, 2001; Brettschneider and Burgess, 2017). We define an episode as all the day-stock information between a buy and a sell date. An episode is classified as a gain if the selling price is higher or equal than the buy price. It is classified as a loss otherwise.

We now introduce a variable: the distance from the extreme at time t, d_t, which we will refer to as "distance". Time t is defined in terms of trading days.

Definition 1. The distance from the extreme is defined as

$$d_t = \begin{cases} \frac{t-t_{max}}{t}, & \text{if episode ends up as a gain} \\ \frac{t-t_{min}}{t}, & \text{if episode ends up as a loss} \end{cases} \tag{1}$$

where t is the number of days since the episode started, and t_{max} and t_{min} are the days when the current maximum and minimum prices of the episodes realized, respectively. Hence,

t is always bigger or equal than t_{max} and t_{min}. These are calculated taking the starting point of an episode equal to $t = 0$.

2.1 Threshold Investigation

We introduce the research question we will address in the corresponding subsection in Section 3.

- *Do investors follow a threshold strategy? That is to say, do they stop on the day when maximum or minimum realizes?*
- *Which categories of investors are more likely to follow a threshold strategy?*

Threshold strategy in the loss domain is not discussed from a theoretical point of view in Strack and Viefers (2019), hence our analysis is more agnostic than the one we make for the threshold strategy in the gain domain. We will assume that a threshold strategy is rational also for the loss domain.

We take into account episodes whose length is shorter or equal to 300 days from buy to sell date (209 trading days). We want to be sure we capture active trading decisions and that we are not looking at decisions of buy-and-hold long term investors (Benartzi and Thaler, 1995; Heath et al., 1999; Brettschneider and Burgess, 2017). We focus on the sample of bank accounts for which demographics are available. The sample is summarized in Table 1. An episode where the investor follows a threshold strategy is characterized by the condition $d_T = 0$ where T is the selling date for that episode. This is the proxy we are going to use to define a threshold strategy.

Definition 2. A trading episode is said to be a threshold strategy episode if $d_T = 0$.

Definition 2 gives a necessary but not sufficient condition to define an investment as a threshold episode. However, since our aim is rejecting a threshold strategy, we are only adding more obstacles to our goal by taking into account a less stringent hypothesis. Being able to reject it would imply, *a fortiori*, that a threshold strategy does not hold.

In Table 1 we see the distribution of some statistics for our sample. We notice that they are in line with the idea that investors suffer from the disposition effect. In particular, d_t is lower for gains than for losses. This is a signal that investors tend to realize gains quicker and have a strong aversion to realize losses at a minimum. Overall, the vast majority of trades does not adhere to a threshold strategy. That is shown in Figure 2. Our definition of threshold strategy applies only to 31.6% of episodes in the gain domain, and to 25.8% of trades in the loss domain.

Table 1: **Summary Statistics for the sample used in the Threshold analysis.** Gain refers to the sample where only investments which resulted in a gain are considered (return higher or equal than 0). Loss refers to the sample where only investments which resulted in a loss are considered. All considers those bank account where at least a gain and a loss were realised. Bank accounts are classified as Cash (the standard one), Keogh or IRA (two types of retirement accounts), Margin or Schwab (two sophisticated products available to investors). Client Segment: Affluent if at any point in time she has more than $\$100,000$ in equity, active if she makes more than 48 trades in any year and General for the residual individuals. If traders could be classified as both affluent and active they were classified as active traders.

	Gain	Loss	All
Percentage of Threshold Episodes	0.316	0.258	0.293
Number Bank Account	15,624	11,390	8,674
Mean Rate of Threshold Consistency per Bank Account	0.275	0.216	0.257
Median Rate of Threshold Consistency per Bank Account	0.043	0	0.250
Mean Number of Episodes per Bank Account	4.640	3.954	11.493
Median Number of Episodes per Bank Account	2	2	6
Number of Cash Bank Accounts	2,591	1,812	1,218
Number of IRA Bank Accounts	2,798	1,674	1,227
Number of Keogh Bank Accounts	91	76	51
Number of Margin Bank Accounts	2,436	1,884	1,469
Number of Schwab Bank Accounts	7,708	5,944	4,709
Number of General Traders	10,368	7,080	5,085
Number of Affluent Traders	2,134	1,549	1,045
Active Trader (num.)	3,122	2,761	2,544
Mean Age	49.70	50.58	50.25
Median Age	48	48	48
Mean Income per Bank Account	6.219	6.224	6.213
Median Income per Bank Account	6	6	6
Number of Females	1,329	949	689
Number of Males	11,947	8,717	6,627
Number of Not Professional Traders	10,362	7,672	5,916
Number of Professional Traders	880	632	462
Number of Traders with Other or unknown Occupation	4,382	3,086	2,296

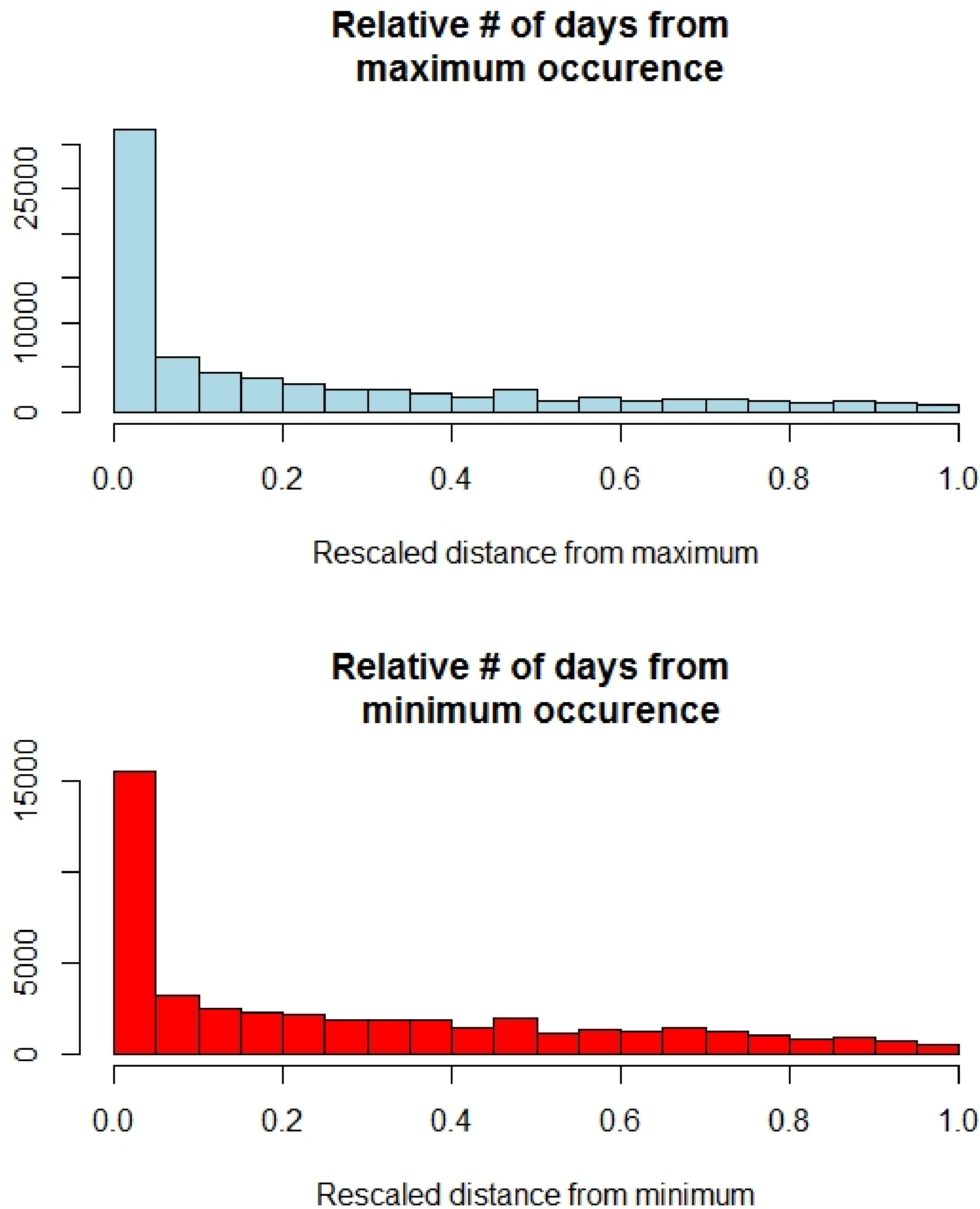

Figure 2: **Rescaled distance in time from the extreme on selling day.** Top: frequency of d_T, the rescaled distance in time from the maximum day calculated on the day when a sell for a gain took place. One observation per episode. Bottom: frequency of d_T, the rescaled distance in time from the minimum day calculated on the day when a sell for a loss took place. One observation per episode.

These are descriptive statistics which clearly hint that a threshold strategy is not consistently followed by our population of investors. In Section 3 we are going to investigate how the propensity to follow a threshold strategy changes at the individual level.

2.2 Maximum Investigation

Our second research question is:

- *How does the propensity to sell a stock vary with respect to the three following variables?*
  - *Level of the price;*
  - *Distance in time from the day of running maximum realisation;*
  - *Distance in price from the running maximum.*

In this section we restrict our attention to a random sample of 13000 episodes. We look at a sample of investments which lasted no longer than 300 days (209 trading days) and resulted in a gain. The choice of restricting the sample to such a period comes from the fact that we want to guarantee the proportional hazard assumption holds and we base our estimation on the idea that investors attention does not span a very long period (Benartzi and Thaler, 1995; Brettschneider and Burgess, 2017). We restrict our attention to the gain domain since it would be difficult estimating the impact of the maximum price on the propensity to sell a stock for a loss, given that the maximum price would often coincide with the purchase price. Hence, it would be difficult to disentangle if a stock is being sold for being far from the past maximum or for being far from the purchase price. On top of that, Strack and Viefers (2019) test their predictions in a laboratory setting where stocks are on average in the gain domain (selling at a loss is a dominated strategy in their experiment). We exclude the 10% of most volatile episodes[1]. Since we are interested in how the propensity to sell a stock changes when the distance of the price from the past maximum increases, we did not want to focus on those trades where the likelihood of big intraday drops in price is high. The characteristics of our sample are summarised in Table 2.

We analyse the data using the proportional hazard (PH) model, developed by Cox (1972). Survival analysis models are widely used in medical research and they are relatively popular in demography and labour economics. They have recently been used in a series of financial applications (Ivković and Weisbenner, 2005; Deville and Riva, 2007; Jiao, 2015; Brettschneider and Burgess, 2017).

[1]We get the average daily ratio of minimum to maximum price per each investment episode and we exclude those trades where the ratio is lower or equal than 0.93

Table 2: **Summary Statistics for the sample used in the Maximum analysis.**

Number Bank Accounts	8704
Number Trading Episodes	13000
Mean Return per Episode	1.19
Median Return per Episode	1.12
Mean Length per Episode (trading days)	68.07
Median Length per Episode (trading days)	52

The proportional hazard model is a semi-parametric model, aimed at describing the "time-to-event" of individuals. In our case the time to event is the time from the start to the end of an investment episode. It has the advantage of assessing the impact of the covariates over the entire time axis, while for example a logistic regression only evaluates the odds of the event/non event with respect to a fixed time. Informally, a PH model incorporates all the information accumulated in time for a given episode. A logistic regression would evaluate the information in a given day independently from the information on other days of the same episode. We now add more details to our discussion. In particular, we need to define some objects

Definition 3. Let T be a non-negative continuous random variable, representing the time until the event of interest.

$F(t) = P(T \leq t)$ denotes the distribution function and $f(t)$ the probability density function of random variable T.

$S(t) = P(T > t) = 1 - F(t)$ is the survival function. It is the probability that a randomly selected individual will survive beyond time t. It is a decreasing function, taking values in $[0, 1]$ and it equals 1 at $t = 0$ and 0 at $t = \infty$.

$\mathrm{H(t)} = -\,\mathrm{logS(t)}$ is the cumulative hazard function.

The hazard function (or hazard rate) $h(t)$ measures the instantaneous risk of dying right after time t given the individual is alive at time t.

In particular, we will fit a regression model where we evaluate the change in the hazard rate with respect to a set of covariates. That is to say, given a set of covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})$ measured for subject i, the following model is fit to the data

$$h_i\left(t\right) = h_0\left(t\right)\exp\left(\beta^t \mathbf{x_i}\right);$$

with β, a $p \times 1$ vector of parameters and $h_0(t)$ which is the baseline hazard function (i.e. hazard for a subject i with $\mathbf{x_i} = \mathbf{0}$).

The proportional hazards assumption states that the ratio of the hazards of two subjects with covariates x_i and $x_{i'}$ is constant over time:

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{\exp(\beta^t \mathbf{x}_i)}{\exp(\beta^t \mathbf{x}_{i'})}$$

The Cox PH model is a semi-parametric model. It means that it leaves the form of $h_0(t)$ completely unspecified and it estimates the model in a semi-parametric way. Then, to estimate the model we maximize a partial likelihood. Finally, we should point out that we are estimating a model with time changing covariates so it is better to define it as

$$h_i(t) = h_0(t) \exp\left(\beta^t \mathbf{x}_{it}\right)$$

To deal with unobserved heterogeneity we stratify the model based on the investor who holds the position. This means that each bank account has a different baseline hazard function, which can absorb any heterogeneity not captured by the model covariates Hence, the hazard function for the i_{th} position of the j_{th} bank account is

$$h_{ij}(t) = h_{0j}(t) \exp\left(\beta^t \mathbf{x}_{ijt}\right);$$

where $\mathbf{x}_{ijt}$ is the covariate vector for the position. A final word should be spent on why we need to stratify at the bank account level. Possible reasons for there being a difference between investors include their preference for risk, their beliefs about the market (e.g. whether there is price momentum or not), their investment objectives and the particular strategy they are following. For example, some investors may trade very frequently and follow a strategy based on short term changes in stock prices. The holding periods of these investors will therefore be shorter than other investors in the sample. Differentiating investors baseline hazards can separate out this kind of differences from the effects of the covariates included in the model, that are common across all investors.

3 Results

3.1 Threshold strategy investigation

We already anticipated that we say that an investment episode respect a threshold strategy if it satisfies Definition 2, hence if the stock was sold on the maximum day for a gain or on the minimum day a loss. We already saw in Section 2 that the vast majority of trading episodes is not consistent with a threshold strategy. The unit of analysis in this section will be the bank account, since we are interested in the rate of consistency with threshold for each bank account. In table 3 we analyse the rate of threshold consistency per bank account. We perform a negative binomial regression. The dependent variable in our negative binomial regression is defined as:

- N_g, the number of investments in a bank account which were realised on the day when the maximum since purchase price realised and resulted in a gain, for columns 1 and 2 of Table 3.
- N_l, the number of investments in a bank account which were realised on the day when the minimum since purchase price realised and resulted in a loss, for columns 3 and 4 of Table 3.
- $N = N_g + N_l$, for columns 5 and 6 of Table 3.

In columns 1 and 2 of Table 3 we only look at those bank accounts where at least an investment which resulted in a gain was recorded. In columns 3 and 4, we look at those bank accounts where at least an investment which resulted in a loss was recorded. In columns 5 and 6, we look at those bank accounts where at least an investment which resulted in a gain and at least one which resulted in a loss were recorded. To control for the possibility that different investors completed different numbers of trading episodes, we take into account the logarithm of the number of completed episodes (completed gains, losses or all depending on the regression) as an offset. Hence, we measure how the rate of threshold consistency varies from one bank account to the other. We now introduce the covariates we are taking into account. They are all defined at the bank account level.

- Dummy for the account type: Cash Account which is a standard bank account, IRA and Keogh, which are two different types of retirement accounts, Margin accounts and Schwab One accounts (more sophisticated products available to the investors);
- Client Segment: Affluent if at any point in time she has more than $\$100,000$ in equity, active if she makes more than 48 trades in any year and General for the residual individuals. If traders could be classified as both affluent and active they were classified as active traders;
- Age in decades;

- Income is classified as a numeric variable which takes values from 1 to 9 and increases with income of the individual[2].
- Gender;
- Occupation, we follow Dhar and Zhu (2006): non-professional if the trader has a "white collar/clerical","blue collar/craftsman" or "service/sales" job; professional occupation if the trader has a "professional/technical" or "administrative/managerial" occupation; the residual category is everyone else[3].

We see that investors with margin accounts and Schwab accounts are more likely to follow a threshold strategy. In particular, investors with margins accounts show a rate of threshold consistency which is, on average, 20% higher than cash accounts both for gains and for losses, separately (almost 30% when we consider the overall rate). Schwab account holders have a consistency rate which is around 10% higher than cash accounts for gains and around 15% for losses. If we believe that a threshold strategy is the rational choice for an investors, we see that more sophisticated investors (those who have margin accounts and Schwab accounts) are more likely to adopt it. This is in line with the idea that sophistication lowers investment biases (Grinblatt and Keloharju, 2001; Dhar and Zhu, 2006). Retirement accounts show a higher rate of threshold consistency than cash accounts but only when the overall rate is considered. Active traders have a higher rate of threshold consistency. That is especially true for losses (6 to 8% higher rate than general traders). That is to be expected since frequency of trading increases the chances that investors are constantly monitoring their investments. Consistency with threshold depends also on attention (Seasholes and Wu, 2007; Barber and Odean, 2008) since investors might lose the possibility to stop at a threshold because of inattention. Affluent traders are less consistent than general traders with a threshold strategy but the effect boils down when we take into account age. That is probably due to the fact that affluent traders are much older, on average, than general traders. Older investors are less likely to be consistent with a threshold strategy. Every ten years, the rate of threshold consistency decreases by around 8%, 4% and 7% in the gain, loss and overall sample. This finding can be linked to the idea that older individuals have lower decision making abilities and make worse financial decisions (Korniotis and Kumar, 2011; Bruine De Bruin, 2017). The higher is the income of the traders, the less likely they are to follow a threshold strategy. Males are more likely to follow a threshold strategy than

[2] 1 corresponds to less than $15,000 per year; 2 to $15,000 to $19,999; 3 to $20,000 to $29,999; 4 to $30,000 to $39,999; 5 to $40,000 to $49,999; 6 to $50,000 to $74,999; 7 to $75,000 to $99,999; 8 to $100,000 to $124,999; 9 to $125,000 or more

[3] Other or unknown occupation

Table 3: **Negative binomial regression for threshold strategy consistency.** Odds ratios with 95% c.i. of a Negative Binomial regression where each observation corresponds to a bank account. The dependent variable is the number of times the investor stopped at a threshold in the gain, loss or overall. There is an offset equal to the number of episodes in the bank account (in the gain, loss, overall).

	Gain		Loss		All	
	(1)	(2)	(3)	(4)	(5)	(6)
Account Type (ref. Cash)						
Account Type IRA	1.068*	1.044	1.097*	1.082	1.105***	1.096**
	(0.997,1.143)	(0.965,1.129)	(0.998,1.207)	(0.971,1.206)	(1.031,1.185)	(1.012,1.187)
Account Type Keogh	1.111	1.180	1.267*	1.144	1.238**	1.262**
	(0.900,1.367)	(0.917,1.516)	(0.980,1.628)	(0.840,1.545)	(1.025,1.494)	(1.014,1.570)
Account Type Margin	1.194***	1.186***	1.237***	1.195***	1.289***	1.275***
	(1.119,1.274)	(1.099,1.280)	(1.135,1.350)	(1.080,1.324)	(1.210,1.373)	(1.184,1.374)
Account Type Schwab	1.129***	1.092***	1.152***	1.130***	1.202***	1.169***
	(1.068,1.194)	(1.023,1.167)	(1.069,1.244)	(1.033,1.236)	(1.137,1.270)	(1.094,1.248)
Client Segment (ref. General)						
Client Segment Affluent	0.905***	0.951	0.871***	0.952	0.863***	0.911**
	(0.851,0.962)	(0.884,1.023)	(0.798,0.950)	(0.860,1.053)	(0.810,0.920)	(0.844,0.981)
Client Segment Active	1.036*	1.092***	1.059**	1.078**	1.076***	1.117***
	(0.998,1.076)	(1.045,1.142)	(1.009,1.111)	(1.017,1.142)	(1.040,1.113)	(1.073,1.163)
Age (decades)		0.921***		0.958***		0.932***
		(0.906,0.936)		(0.938,0.978)		(0.919,0.946)
Income		0.990*		0.977***		0.985***
		(0.979,1.000)		(0.964,0.991)		(0.976,0.995)
Male		1.008		1.141**		1.058
		(0.931,1.093)		(1.021,1.277)		(0.979,1.144)
Occupation (ref. Other or unknown)						
Non Professional Occupation		1.060		1.004		1.031
		(0.977,1.150)		(0.898,1.121)		(0.954,1.113)
Professional Occupation		1.015		0.956		0.994
		(0.971,1.061)		(0.901,1.014)		(0.954,1.036)
McFadden Adj. R^2	0.24	0.46	0.25	0.46	0.25	0.46
Observations	15,624	11,477	11,390	8,315	8,674	6,280

Note: *p<0.1; **p<0.05; ***p<0.01

females for losses. Their rate of threshold consistency is 14.1% higher. We do not see any differences based on the occupation of the traders.

The take home from this section is that investors do not consistently follow a threshold strategy, in line with Strack and Viefers (2019). The main differences in threshold consistency are due to age, client segment and account type. Sophisticated investors and active traders are more consistent with a threshold strategy. Affluent and older investors are less consistent than general investors with a threshold strategy. Males are more willing to realise losses at a threshold than females.

3.2 Maximum price investigation

We can now focus on the main results of our work. First, we introduce the variables of interest. They are measured at the daily level. The buy date of any investment episode is date 0. Every date is registered as the difference in trading days between that date and the buy date and denoted by t. Only the propensity to sell on days when the stock is trading above the buy price is estimated. That means that the information on those days in which the stock is trading at a loss is not incorporated in the estimate. We confine ourselves to estimate the propensity to sell for a gain. Hence, we thought it was not appropriate estimating the propensity to sell the stock on those days when it was trading at loss, since we constrained it to zero. To make a parallel with the medical literature, from which we borrow our estimation strategy, think about an allergy which we know can only occur during the spring. It would not make sense estimating the probability of occurrence based on the covariates measured during the winter. We define now our covariates of interest, measured at any given day t.

- Ratio to Max Price (Ratiomax) is the ratio of the daily closing price to the maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of the selling price to the maximum price in the episode. We split it into quartiles based on the stock-bank account-day distribution. Defined as low in the interval [0.349; 0.918]; medium-low in the interval (0.918; 0.957]; medium-high in the interval (0.957; 0.981] and high in the interval (0.981; 1].

- Distance is the rescaled distance from the occurrence of the running maximum date, as we defined it in definition 1, $\frac{t-t_{max}}{t}$. t_{max} is the day when maximum price between day 0 and day t realised. We split it into tertiles based on the stock-bank account-day distribution. Distance is defined as low in the interval [0; 0.07]; medium in the interval [0.07; 0.34) and high in the interval [0.34; 1];

- Return is the ratio of the daily closing price to the purchase price in the investment episode. On the selling date it is equal to the ratio of the selling price to the purchase price in the episode. We split it into tertiles based on the stock-bank account-day distribution. Defined as low in the interval [0.58; 1.01]; medium in the interval (1.06; 1.17]; high in the interval (1.17, 5.53].

We highlight that they are not fixed at the investment episode level, they are updated from day to day for any investment. We make an example for the Distance to clarify the point. Let's say we are observing an episode on day 10 and the maximum price since purchase occurred on day 5. On day 10, the value of Distance will be $\frac{10-5}{10} = \frac{1}{2}$ and on day 11 it will be $\frac{11-5}{11} = \frac{6}{11}$. Then, if the maximum price on day 12 is higher that the maximum price registered on day 5, Distance on day 12 will be equal to $\frac{12-12}{12} = 0$ and so on, so forth. We rescale all the variables since we need consistency from one trading episode to another. Prices are really different from stock to stock. On top of that, we use the categorical version of the variables instead of using their continuous version for two reason. First, to take into account non linearity. Second, to have a model which could be analysed through proportional hazard technique. In our specification, the proportional hazard assumption is not violated for any of the models. On top of that, we are able to capture the main non-linear changes in the effect of the variables. In the appendix we repeat the analysis referring to the absolute distance in trading days from the maximum day. The most important implication of the proportional hazard assumption is that the effect of a covariate is constant in time. Whilst a violation of the assumption does not invalidate the model, it does significantly alter the interpretation. If an effect does change over time then the hazard ratio is only an average of this process, and if it changes a lot then this average can be misleading. When we take into account the absolute distance in trading days from the maximum day (see the appendix), the proportional hazard assumption no longer holds. That means that the effect of being one day after the maximum day has a different impact on the propensity to sell a stock, if the effect is evaluated for example on day 7 or on day 30 since purchase. That means that rescaling the variables was necessary. In particular, we find that the propensity to sell is the highest on the maximum day and up to two days after the maximum occurred, while it declines from the third day onwards (see the appendix)

In Table 4 we estimate three proportional hazard models where we take into account the effect of Ratiomax, Distance and Return. We stratify the baseline hazard at the bank account level, in order to take into account differences in the propensity to realise a stock due to fixed investors' characteristics and we control for time effects (month and year). We see that the propensity to sell is lowest when the stock is trading close to the maximum price. The probability of selling at a high Ratiomax is 28% lower than the probability of

Table 4: **Proportional Hazard model of the hazard of selling a stock. Single covariates analysis.** The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Odds Ratios with c.i. Baseline hazard rate stratified at investor level. Clustered robust s.e. at bank account level. Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. Distance in Time from Max Day is the rescaled distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1]. Return is the ratio of daily closing price to the purchase price in the investment episode. On the selling date it is equal to the ratio of selling price to the purchase price in the episode. Low [1; 1.06]; Medium (1.06; 1.17]; High (1.17, 5.53].

	(1)	(2)	(3)
Ratio Price to Max Price (ref. Low)			
Medium-Low	0.909		
	(0.792,1.043)		
Medium-High	1.062		
	(0.932,1.210)		
High	0.720***		
	(0.619,0.837)		
Dist. in Time from Max Day (ref. Low)			
Medium		0.877**	
		(0.786,0.979)	
High		0.430***	
		(0.385,0.481)	
Return (ref. Low)			
Medium			2.719***
			(2.435,3.035)
High			3.435***
			(2.988,3.949)
Xu-O'Quigley R^2	0.020	0.061	0.10
Concordance	0.57	0.61	0.64
PH Assumption Valid (0.05)	YES	YES	YES
Time Controls	YES	YES	YES
Number of Trading Episodes	13,000	13,000	13,000
Number of Bank Accounts	8,704	8,704	8,704
Observations	621,849	621,849	621,849
Note:			*p<0.1; **p<0.05; ***p<0.01

selling at other points. This contradicts the predictions of dynamic regret, when the agent is only focused on regret about past decisions. Propensity to sell peaks at a medium-high level of the Ratiomax but differences among low, medium-low and medium-high categories are not significant. We conclude that regret does not bite as we expected. Propensity to sell is the highest when the price is close but not extremely close to the maximum. The pattern for Distance is quite strong and clear. Propensity to sell is 12.3% lower at a medium with respect to a low Distance and 57% lower at high distance. To reconcile these findings with what we mentioned before referring to the absolute distance in days (see the appendix for a more detailed discussion), we notice that more than half of the times that the rescaled distance is low, it means that the distance from the past maximum is not greater than 2 days. We conclude that the propensity to sell a stock peaks in those few days around maximum realisation. We suggest that regret about past forgone opportunities, might be a channel to explain this finding. It is a different form of regret from the one discussed in Strack and Viefers (2019). They only discuss the regret coming from selling at a price lower than the past maximum and they do not discuss how the distance in time from the maximum realisation shapes the propensity to sell. Here, we suggest that investors might experience regret from selling at a time further from the time when the maximum realised. Hence, the propensity to sell on a day further in time from the past maximum is lower because the regret experienced by the investor is higher. Higher returns increase the propensity to sell steadily. When returns are in the medium or high region (above 6% return) the rate of selling is around 3 times the rate of selling when the stock is in the Low return region (below 6%). This confirms the predictions of Strack and Viefers (2019) and the evidence in Ben-David and Hirshleifer (2012).

Since the effect of Distance is very relevant in terms of both strength and explanatory power, we are interested in the interaction of it with the effect of Ratiomax. In the appendix we report the joint distribution of the two categories. At least 2% of stock-bank account-day observations fall in each category. From Figure 3 we can see that the interaction between Ratiomax and Distance leads to an interesting scenario. When the Distance is low or medium, the average percentage of selling days is the highest when Ratiomax is low. There is probably a panic effect, which can be framed as regret but is much deeper than what we described before. Investors are more willing to realise a gain when it is closest in time to maximum but furthest in price. If regret has role, it is through this channel. Investors regret selling at a time far from maximum. However, they sell as soon as the price decreases significantly. We can see this by looking at the fact that the percentage of sale days is 2.5% at high Ratiomax and low Distance and it is 4% at low price and low distance. Hence, it looks like anticipated regret of incurring higher losses is higher than the experienced regret given by the distance

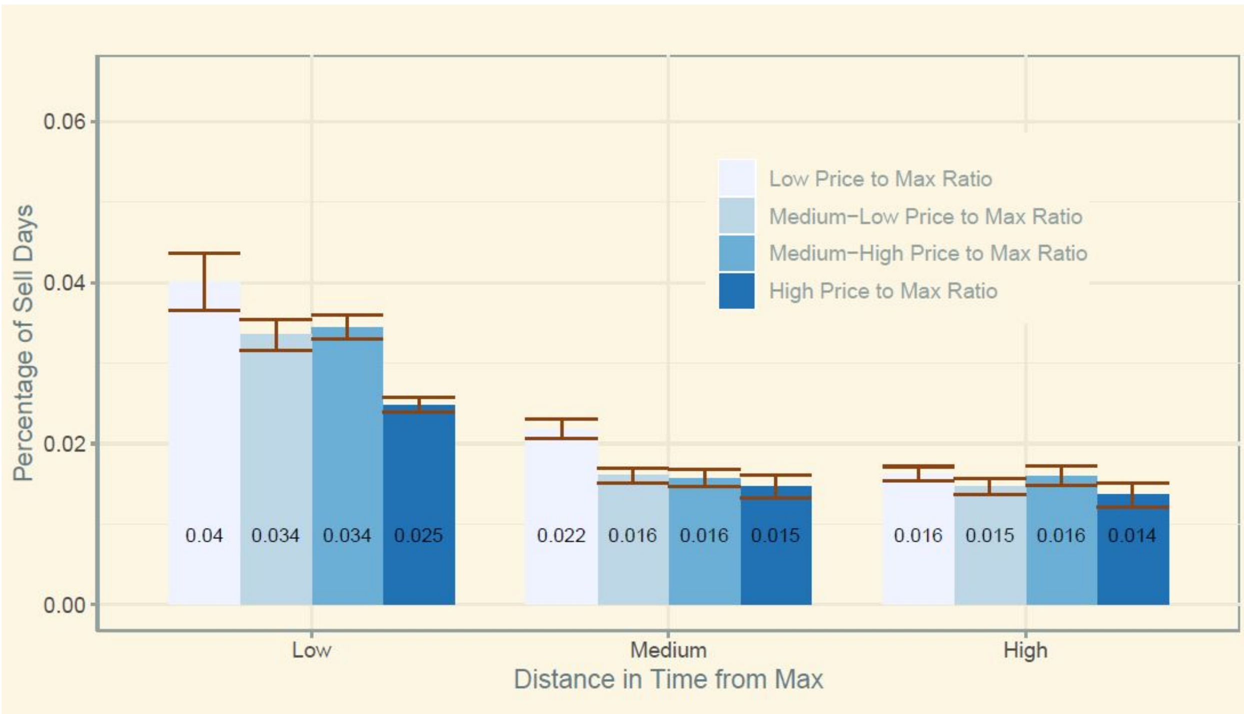

Figure 3: **Percentage of selling days for each combination of RatioMax and Distance.** Percentage of selling days out of all stock-bank account-days per each category (0.95 c.i.). Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. Distance in Time from Max Day is the rescaled distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1].

in price from the past maximum. This is in line with the evidence of anticipated regret of Fioretti et al. (2018).

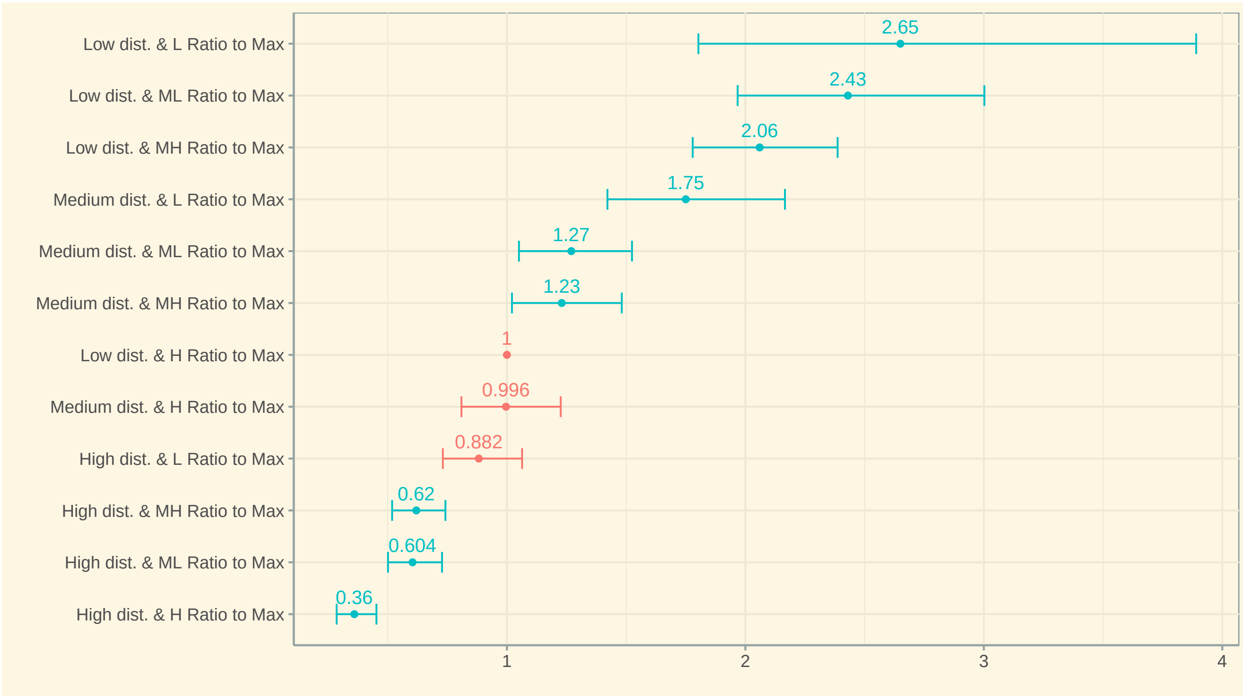

Figure 4: **Odds Ratio Proportional Hazard model: RatioMax and Distance.** Odds Ratio from a Proportional Hazard model of the hazard of selling a stock. Each Odds Ratio corresponds to the effect of being in a given category. The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Clustered robust s.e. at bank account level. Blue odds ratios are significantly different from 1 ($p < 0.05$), red are not. Proportional Hazard assumption valid at 0.05 level. Xu-O'Quigley R^2 is 0.095. Concordance is 0.65. Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. Distance in Time from Max Day is the rescaled distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1]. The baseline category is "Low Distance and High RatioMax".

We fit a proportional hazard model where we take into account all the 12 combinations of Ratiomax and Distance categories. All categories in our regression are disjoint. The regression table is reported in the appendix. In Figure 4 we report the odds ratio of the 11 coefficients. The baseline category corresponds to "Low Distance and High Ratiomax", the ideal point to sell a stock, from an accounting perspective. We can see that all cases where Distance is low are clustered at the top, when we rank categories based on the propensity to sell for each of them. All cases where Distance is high are clustered at the bottom. The big exception is the baseline category, low Distance and high Ratiomax. Figure 4 offers a more insightful representation with respect to the sample averages we reported in Figure 3. The hazard of selling is 2.65 higher for low distance and low Ratiomax stock-days than at the baseline. In general, when the stock is close in time but not close in price (i.e. low distance but not high Ratiomax) the hazard of selling is always estimated to be at least double than that at the baseline. The complete picture seems to suggest that reality is more complicated than how the lab describes it. Distance in price from the maximum plays a role but the effect is not nice and linear as the one observed by Strack and Viefers (2019). When the distance in price is considered in isolation, it looks like the propensity to sell peaks at a point which is close but not the closest possible to the maximum. We can safely claim that traders are more willing to sell at a low distance in time from the maximum. They probably wait for a new maximum when a lot of time has passed since the last one. Hence, it is always a salient figure in their mind. However, when the distance in time is short, they are more willing to sell stocks which are further from the past maximum. We believe that panic might play a big role. It looks like investors are not extremely good at catching the best time to realise a stock and decide to realise it only when the price path shows a defined descending trend. Predictions of regret theory in dynamic context by Strack and Viefers (2019) are only partially confirmed then.

4 Conclusion

In our work we tested if investors stop at an optimal ex ante threshold and are prone to regret in their decision to sell stocks. We were motivated by the theoretical and experimental evidence of Strack and Viefers (2019). A threshold strategy implies that an investor never sells a stock at a price where she previously decided not to sell. We rejected that hypothesis for our sample of investors. We further investigated investors differences in the propensity to adopt a threshold strategy. First, we saw that investors are more willing to adopt a threshold strategy in the gain with respect to the loss domain. That is a consequence of the disposition effect, the higher propensity to realise gains with respect to losses (Shefrin and Statman, 1985; Odean, 1998; Barber and Odean, 2013). Second, we found that the main differences in the rate of threshold consistency are due to age, client segment and account type. Sophisticated investors and active traders are more consistent with a threshold strategy. Affluent and older investors are less consistent than general investors with a threshold strategy. Males are more willing to accept losses at a minimum. Affluent and older investors are more likely to depart from a rational threshold strategy, this is linked to the idea that decision making of older individuals is poorer (Korniotis and Kumar, 2011; Bruine De Bruin, 2017).

We investigated the impact of the running maximum price in an investment episode on the propensity of investors to realise gains. We fitted a proportional hazard model to the propensity to sell for a gain. Predictions of regret theory in a dynamic context (Strack and Viefers, 2019) are that the propensity to sell increases with the level of the price and decreases with the distance of the price from the past maximum. The first prediction was confirmed, since the propensity to sell a gain strongly increases as the return increases (consistent with Ben-David and Hirshleifer, 2012). The effect of the distance in price from the past maximum is not the one predicted by Strack and Viefers (2019). We did find that the propensity to sell is actually lower when the stock is trading very close to the maximum price, while it is highest when the price is close to the past maximum but not in the closest region.

We also investigated the impact that the distance in time from the past maximum has on the propensity to sell a stock. We found a very strong effect, with the propensity to sell a stock falling as the distance in time from the past maximum increases. We suggest this might be a form of regret not considered by Strack and Viefers (2019). Investors might experience more regret the further they are in time from the day when the maximum realised. This regret might reduce their propensity to sell. On top of that, we investigated the joint effect of distance in time and distance in price from the past maximum, finding that investors are more willing to realise stocks which are closer in time but further in price from the past maximum. When the time distance from the past maximum is short, the predictions of regret reverse. Investors are more willing to realise stocks, the further is the price from the maximum. Two forces are at work: investors are willing to wait for a new maximum to occur if a long time has passed since the last one and they panic when the stock price drops a short time after the maximum realised. Anticipated regret (Fioretti et al., 2018) is a possible explanation. Investors might be focused on the possibility that the stock price might decrease even more and they rush to sell to minimise regret which has not materialised yet.

References

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*(2), 773–806 (cit. on p. 4).

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, *116*(1), 261–292 (cit. on p. 4).

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818 (cit. on pp. 3, 12).

Barber, B. M., & Odean, T. (2013). The behavior of individual investors (G. M. Constantinides, M. Harris, & R. M. Stulz, Eds.). In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.). Elsevier. (Cit. on pp. 1, 20).

Barberis, N., Huang, M., & Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, *96*(4), 1069–1090 (cit. on p. 3).

Baucells, M., Weber, M., & Welfens, F. (2011). Reference-point formation and updating. *Management Science*, *57*(3), 506–519 (cit. on p. 4).

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, *110*(1), 73–92 (cit. on pp. 5, 8).

Ben-David, I., & Hirshleifer, D. (2012). Are investors really reluctant to realize their losses? trading responses to past returns and the disposition effect. *The Review of Financial Studies*, *25*(8), 2485–2532 (cit. on pp. 4, 17, 21).

Bleichrodt, H., & Wakker, P. P. (2015). Regret theory: A bold alternative to the alternatives. *The Economic Journal*, *125*(583), 493–532 (cit. on p. 1).

Braun, M., & Muermann, A. (2004). The impact of regret on the demand for insurance. *Journal of Risk & Insurance*, *71*(4), 737–767 (cit. on p. 3).

Brettschneider, J., & Burgess, M. (2017). Using a frailty model to measure the effect of covariates on the disposition effect. *CRISM Working Paper Series*, *17*(5) (cit. on pp. 4, 5, 8).

Bruine De Bruin, W. (2017). Ageing and economic decision-making. In B. Ranyard (Ed.), *Economic psychology* (pp. 371–386). John Wiley & Sons, Ltd. (Cit. on pp. 12, 20).

Coval, J. D., Hirshleifer, D. A., & Shumway, T. (2005). Can individual investors beat the market? *HBS Finance Working Paper*, *34* (cit. on p. 4).

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220 (cit. on pp. 1, 8).

Descamps, A., Massoni, S., & Page, L. (2016). Knowing when to stop and make a choice, an experiment on optimal sequential sampling. *Available at SSRN 2883847* (cit. on p. 3).

Deville, L., & Riva, F. (2007). Liquidity and arbitrage in options markets: A survival analysis approach. *Review of Finance*, *11*(3), 497–525 (cit. on p. 8).

Dhar, R., & Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, *52*(5), 726–740 (cit. on pp. 1, 12).

Fioretti, M., Vostroknutov, A., & Coricelli, G. (2018). Dynamic regret avoidance. *USC-INET Research Paper* (cit. on pp. 3, 19, 21).

Firth, D. (2003). Overcoming the reference category problem in the presentation of statistical models. *Sociological Methodology*, *33*(1), 1–18 (cit. on pp. 26, 31).

Firth, D., & De Menezes, R. X. (2004). Quasi-variances. *Biometrika*, *91*(1), 65–80 (cit. on pp. 26, 31).

Gollier, C., & Salanié, B. (2006). Individual decisions under risk, risk sharing and asset prices with regret. *Working paper, University of Toulouse* (cit. on p. 3).

Grinblatt, M., & Keloharju, M. (2001). What makes investors trade? *The Journal of Finance*, *56*(2), 589–616 (cit. on pp. 3, 12).

Heath, C., Huddart, S., & Lang, M. (1999). Psychological factors and stock option exercise. *The Quarterly Journal of Economics*, *114*(2), 601–627 (cit. on pp. 3, 5).

Huang, X. (2019). Mark twain's cat: Investment experience, categorical thinking, and stock selection. *Journal of Financial Economics*, *131*(2), 404–432 (cit. on p. 4).

Huddart, S., Lang, M., & Yetman, M. H. (2009). Volume and price patterns around a stock's 52-week highs and lows: Theory and evidence. *Management Science*, *55*(1), 16–31 (cit. on p. 3).

Ivković, Z., Poterba, J., & Weisbenner, S. (2005). Tax-motivated trading by individual investors. *The American Economic Review*, *95*(5), 1605–1630 (cit. on p. 4).

Ivković, Z., Sialm, C., & Weisbenner, S. (2008). Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis*, *43*(3), 613–655 (cit. on p. 4).

Ivković, Z., & Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, *60*(1), 267–306 (cit. on pp. 4, 8).

Jiao, P. (2015). Losing from naive reinforcement learning: A survival analysis of individual repurchase decisions. *Available at SSRN 2574141* (cit. on p. 8).

Kent, J. T., & O'Quigley, J. (1988). Measures of dependence for censored survival data. *Biometrika*, *75*(3), 525–534 (cit. on pp. 25, 26).

Korniotis, G. M., & Kumar, A. (2011). Do older investors make better investment decisions? *The Review of Economics and Statistics*, *93*(1), 244–265 (cit. on pp. 1, 4, 12, 20).

Korniotis, G. M., & Kumar, A. (2013). Do portfolio distortions reflect superior information or psychological biases? *Journal of Financial and Quantitative Analysis*, *48*(1), 1–45 (cit. on pp. 1, 4).

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824 (cit. on p. 1).

Michenaud, S., & Solnik, B. (2008). Applying regret theory to investment choices: Currency hedging decisions. *Journal of International Money and Finance*, *27*(5), 677–694 (cit. on p. 3).

Muermann, A., Mitchell, O., & Volkman, J. (2006). Regret, portfolio choice, and guarantees in defined contribution schemes. *Insurance: Mathematics and Economics*, *39*(2), 219–229 (cit. on p. 3).

Muermann, A., & Volkman, J. M. (2006). Regret, pride, and the disposition effect. *PARC Working Paper Series* (cit. on p. 3).

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, *53*(5), 1775–1798 (cit. on pp. 1, 20).

Odean, T. (1999). Do investors trade too much? *American Economic Review*, *89*(5), 1279–1298 (cit. on pp. 1, 4).

O'Quigley, J., Xu, R., & Stare, J. (2005). Explained randomness in proportional hazards models. *Statistics in Medicine*, *24*(3), 479–489 (cit. on p. 26).

Seasholes, M. S., & Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance*, *14*(5), 590–610 (cit. on p. 12).

Seasholes, M. S., & Zhu, N. (2010). Individual investors and local bias. *The Journal of Finance*, *65*(5), 1987–2010 (cit. on pp. 1, 4).

Shapira, Z., & Venezia, I. (2001). Patterns of behavior of professionally managed and independent investors. *Journal of Banking & Finance*, *25*(8), 1573–1587 (cit. on pp. 1, 4).

Shefrin, H., & 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 (cit. on pp. 1, 20).

Strack, P., & Viefers, P. (2019). Too proud to stop: Regret in dynamic decisions. *Journal of the European Economic Association* (cit. on pp. 1–3, 5, 8, 14, 17, 20, 21, 27, 35).

Xu, R., & O'Quigley, J. (1999). A r2 type measure of dependence for proportional hazards models. *Journal of Nonparametric Statistics*, *12*(1), 83–107 (cit. on pp. 25, 26).

Appendix

Negative Binomial Regression

The Poisson distribution may be generalized by including a gamma noise variable which has a mean of 1 and a scale parameter of ν. The Poisson-gamma mixture (negative binomial) distribution that results is

$$P(Y = y_i|\mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{y_i} \tag{2}$$

where $\mu_i = t_i * \mu$ and $\alpha = \nu^{-1}$. The parameter μ is the mean incidence rate of y per unit of exposure. Exposure may be time, space, distance, area, volume, or population size. Because exposure is often a period of time, we use the symbol t_i to represent the exposure for a particular observation. When no exposure is given, it is assumed to be one. The parameter μ may be interpreted as the risk of a new occurrence of the event during a specified exposure period, t.

In our case the exposure time t_i is the number of episodes in a bank account and μ_i is the number of threshold episodes in the same bank account. In negative binomial regression, the mean of y is determined by the exposure time t and a set of k regressor variables (the x's). The expression relating these quantities is

$$\mu_i = \exp(\log(t_i) + \beta_1 x_{1i} + ... + \beta_k x_{ki}) \tag{3}$$

often $x_1 = \mathbf{1}$, in which case β_1 is called the intercept. We estimate the vector of β coefficients through maximum likelihood.

Pseudo R squared for Cox models

Here, we spell out some details of the pseudo R squared proposed by Xu and O'Quigley (1999), which we report for our proportional hazard models. They start from a coefficient of explained randomness derived by Kent and O'Quigley (1988). The coefficient aims at explaining the variability on the outcome looking at the distribution of time to events, given covariates. That coefficient has the following properties:

- When a covariate is unrelated to survival, and the corresponding regression coefficient is equal to zero, it is equal to zero;

- When the effect of at least a coefficient is different from 0, it is between 0 and 1;

- It is invariant under linear transformations of the covariates and under monotone increasing transformations of time.

The coefficient we use (Xu and O'Quigley, 1999), uses the same basic ideas but looking at the distribution of the covariates at each time. The construction can be carried out using routine quantities calculated during a standard proportional hazards analysis. Inference is also greatly simplified. Most importantly, the presence of time-dependent covariates presents no difficulties for Xu and O'Quigley (1999) coefficient estimation, while the one by Kent and O'Quigley (1988) is not defined in that case. In O'Quigley et al. (2005) there are further discussions on the robustness of the pseudo R-squared we use.

Supplementary Analysis

Here, we include some extra analysis to complement the analysis in the main text. In Table 5 we present the same model we presented in Table 3, taking into account knowledge and experience of investors. As you can see, those figures are available for a relatively small subset of individuals. The sole novelty is represented by the effect of knowledge. Investors with extensive knowledge of investments have a rate of threshold consistency which is 12.7% lower than others, in the gain domain. That might be an effect of overconfidence.

As we mentioned in the text, we report the joint distribution of Ratiomax and Distance (Table 6) and the regression table which corresponds to Figure 4 (Table 7). Moreover, we reproduce Figure 4 estimating confidence intervals using the method of Quasi-variance developed by Firth (2003), Firth and De Menezes (2004). In this way, we overcome the problem of having a reference category for which confidence intervals cannot be estimated and we enhance the possibility of reproducing our work. Figures 6 and 7 show the odds ratios derived from a PH model where we split the data into bins based on the combination of RatioMax and Return (Figure 6) and the combination of Distance and Return (Figure 7. You can see that the propensity to realise a stock is highest for high return levels but not in the region closest to the past maximum. Distance plays an important role, since investors' propensity to sell at a low distance in time from the past maximum is always higher than the propensity to sell at a high distance, for any level of the return. In Table 8 we reproduce the analysis of column 2 in Table 4, using absolute distance in trading days from the maximum instead of the standardized distance. We see that the proportional hazard assumption is not valid for the regression in Table 8. That means that the effect we report is not constant in time. Hence, the hazard ratio we report is only an average effect. That is to be expected since we believe that being 1 or 2 days far from the maximum is not the same when an investor is considering a 7 days or a 50 day investment, intuitively. This backs up the idea that a rescaled distance, like the one we defined in Definition 1 is appropriate. We again see that, on average, the propensity to sell is highest for small distance in time from the maximum day. The absolute distance in time from maximum realisation is a more intuitive measure than the rescaled distance we defined in Definition 1 but it has a much lower explanatory power for our variable of interest. Given this premise, we observe that the propensity to sell a stock is very high on the maximum day and on the two days after. More than one week (5 trading days) after the maximum realised, the propensity to sell a stock is almost 60% lower than on the maximum day. In a range between 3 to 5 trading days from maximum, the propensity to sell is 25.8% lower than on the maximum day, while there are no significant differences in the propensity to sell on the maximum day and on the two days after that. Although the difference is not significant, it is interesting to see that the propensity to sell is highest on the day after the maximum, when it is 9.3% higher than on the maximum day. Hence, the propensity to sell peaks the day after the investor missed the chance to sell at a maximum. This leads to two possible interpretations. The first one is that investors wait until the maximum unfolds and start selling in a time span which is close to it. This is an equivalent mechanism to the one proposed by Strack and Viefers (2019) for price. Regret is lower when the time distance is lower. Another option is attention. When the stock peaks the investor starts paying attention to it and sells it shortly after. However, given that the PH assumption does not hold and the effects are small and not significant, we do not want to push the interpretation too far. Finally, Figure 8 reports Figure 4 from Strack and Viefers (2019).

Table 5: **Negative binomial regression for threshold strategy consistency (extended).** Odds ratios with 95% c.i. of a Negative Binomial regression where each observation corresponds to a bank account. The dependent variable is the number of times the investor stopped at a threshold in the gain, loss or overall. There is an offset equal to the number of episodes in the bank account (in the gain, loss, overall).

	Gain		Loss		All	
	(1)	(2)	(3)	(4)	(5)	(6)
Account Type (ref. Cash)						
Account Type IRA	1.032	1.029	1.081	1.081	1.076	1.068
	(0.904,1.178)	(0.905,1.172)	(0.910,1.286)	(0.914,1.281)	(0.949,1.220)	(0.946,1.207)
Account Type Keogh	1.182	1.212	1.261	1.219	1.255	1.266*
	(0.841,1.654)	(0.872,1.679)	(0.850,1.849)	(0.822,1.783)	(0.946,1.663)	(0.961,1.664)
Account Type Margin	1.219***	1.244***	1.140	1.182*	1.272***	1.292***
	(1.067,1.394)	(1.092,1.418)	(0.956,1.362)	(0.996,1.405)	(1.123,1.442)	(1.146,1.459)
Account Type Schwab	1.162**	1.154**	1.143*	1.160**	1.221***	1.216***
	(1.038,1.304)	(1.032,1.292)	(0.987,1.328)	(1.005,1.343)	(1.097,1.360)	(1.097,1.351)
Client Segment (ref. General)						
Client Segment Affluent	1.004	0.984	0.964	0.970	0.944	0.929
	(0.906,1.111)	(0.889,1.087)	(0.834,1.111)	(0.841,1.115)	(0.850,1.047)	(0.838,1.028)
Client Segment Active	1.081**	1.082**	1.080*	1.097**	1.115***	1.110***
	(1.008,1.158)	(1.012,1.158)	(0.987,1.182)	(1.005,1.197)	(1.049,1.186)	(1.045,1.178)
Age (decades)	0.907***	0.909***	0.931***	0.943***	0.910***	0.917***
	(0.883,0.931)	(0.886,0.932)	(0.900,0.962)	(0.914,0.973)	(0.889,0.931)	(0.897,0.938)
Income	0.994	0.992	0.976**	0.975**	0.989	0.987*
	(0.977,1.011)	(0.976,1.008)	(0.955,0.998)	(0.955,0.996)	(0.975,1.004)	(0.973,1.002)
Male	0.975	0.968	1.083	1.129	0.975	0.989
	(0.858,1.109)	(0.854,1.099)	(0.912,1.292)	(0.954,1.344)	(0.866,1.098)	(0.882,1.111)
Occupation (ref. Other or unknown)						
Non Professional Occupation	1.044	1.051	0.995	1.042	1.010	1.035
	(0.909,1.196)	(0.919,1.198)	(0.822,1.196)	(0.870,1.241)	(0.887,1.147)	(0.915,1.169)
Professional Occupation	1.010	1.002	0.916*	0.927*	0.975	0.973
	(0.945,1.079)	(0.939,1.070)	(0.839,0.999)	(0.852,1.010)	(0.919,1.034)	(0.918,1.031)
Experience (ref. Good)						
Experience Extensive	0.926*		1.041		0.979	
	(0.853,1.005)		(0.940,1.151)		(0.913,1.049)	
Experience Low	1.008		1.044		1.033	
	(0.935,1.086)		(0.944,1.154)		(0.964,1.107)	
Experience None	1.066		0.835		0.973	
	(0.881,1.285)		(0.628,1.095)		(0.815,1.157)	
Knowledge (ref. Good)						
Knowledge Extensive		0.873***		1.055		0.964
		(0.798,0.955)		(0.947,1.175)		(0.895,1.039)
Knowledge Low		1.006		1.012		1.028
		(0.934,1.084)		(0.916,1.118)		(0.961,1.100)
Knowledge None		1.039		1.114		1.079
		(0.932,1.158)		(0.967,1.281)		(0.980,1.188)
McFadden Adj. R^2	0.77	0.76	0.76	0.75	0.76	0.77
Observations	4,713	4,911	3,531	3,663	2,701	2,809
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 6: **Percentage of observations falling in any combination of RatioMax and Distance.** Percentage of stock-bank account-days falling in each category. Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. Distance from Max Day is the rescaled time distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1].

	Low Distance from Max	Medium Distance from Max	High Distance from Max
Low Price Ratio to Max	0.02	0.10	0.13
Medium-Low Price Ratio to Max	0.05	0.10	0.10
Medium-High Price Ratio to Max	0.09	0.09	0.07
High Price Ratio to Max	0.17	0.04	0.04

Table 7: **Odds Ratio Proportional Hazard model: RatioMax and Distance.** Proportional Hazard model of the hazard of selling a stock. The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Odds Ratios with c.i. Baseline hazard rate stratified at investor level. Clustered robust s.e. at bank account level. Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. Distance in Time from Max Day is the rescaled distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1].

Dist. in time from Max and Ratio to Max (ref. Low and High)	
Low dist. and Low Ratio to Max	2.649***
	(1.803,3.891)
Medium dist. and Low Ratio to Max	1.755***
	(1.421,2.166)
High dist. and Low Ratio to Max	0.882
	(0.731,1.064)
Low dist. and Medium-Low Ratio to Max	2.430***
	(1.968,3.002)
Medium dist. and Medium-Low Ratio to Max	1.266**
	(1.051,1.524)
High dist. and Medium-Low Ratio to Max	0.604***
	(0.501,0.728)
Low dist. and Medium-High Ratio to Max	2.061***
	(1.779,2.387)
Medium dist. and Medium-High Ratio to Max	1.230**
	(1.021,1.482)
High dist. and Medium-High Ratio to Max	0.620***
	(0.519,0.742)
Medium dist. and High Ratio to Max	0.996
	(0.809,1.226)
High dist. and High Ratio to Max	0.360***
	(0.286,0.453)
Xu-O'Quigley R^2	0.095
Concordance	0.65
PH Assumption Valid (0.05)	YES
Time Controls	YES
Number of Trading Episodes	13,000
Number of Bank Accounts	8,704
Observations	621,849
Note:	*p<0.1; **p<0.05; ***p<0.01

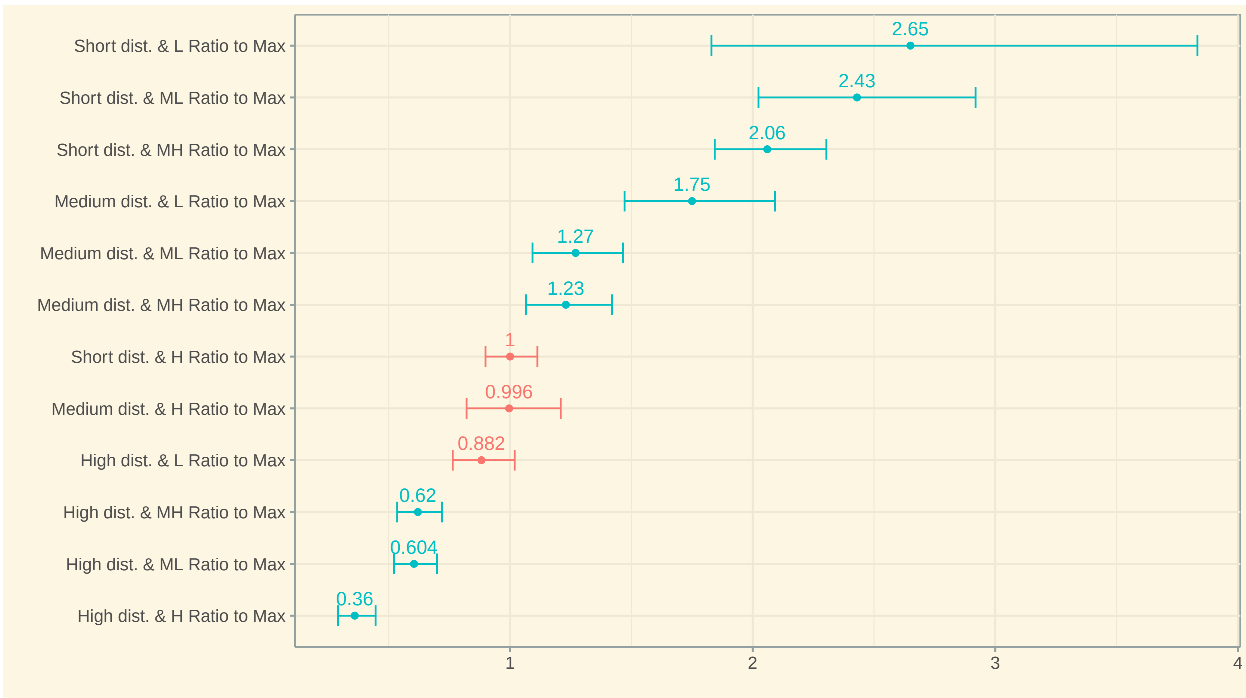

Figure 5: **Odds Ratio Proportional Hazard model: RatioMax and Distance. Firth and De Menezes (2004) method.** Odds Ratio from a Proportional Hazard model of the hazard of selling a stock. Each Odds Ratio corresponds to the effect of being in a given category. The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Clustered robust s.e. at bank account level. Blue odds ratios are significantly different from 1 ($p < 0.05$) when relying upon exact standard errors, red are not. Confidence intervals were obtained with the Quasi-Variance method of Firth (2003), Firth and De Menezes (2004). Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. Distance in Time from Max Day is the rescaled distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1]. The baseline category is “Low Distance and High RatioMax”.

Table 8: **Proportional Hazard model of the hazard of selling a stock. Absolute time distance.** Proportional Hazard model of the hazard of selling a stock. The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Odds Ratios with c.i. Baseline hazard rate stratified at investor level. Clustered robust s.e. at bank account level. Distance in Time from Max Day is measured in trading days.

Dist. from Maximum Day (ref. Max Day)	
1 Day	1.093
	(0.943,1.266)
2 Days	0.929
	(0.783,1.101)
3 to 5 Days	0.742***
	(0.640,0.860)
More than 5 Days	0.413***
	(0.363,0.470)
Xu-O'Quigley R^2	0.061
Concordance	0.61
PH Assumption Valid (0.01)	NO
Time Controls	YES
Number of Trading Episodes	13,000
Number of Bank Accounts	8,704
Observations	621,849
Note:	*p<0.1; **p<0.05; ***p<0.01

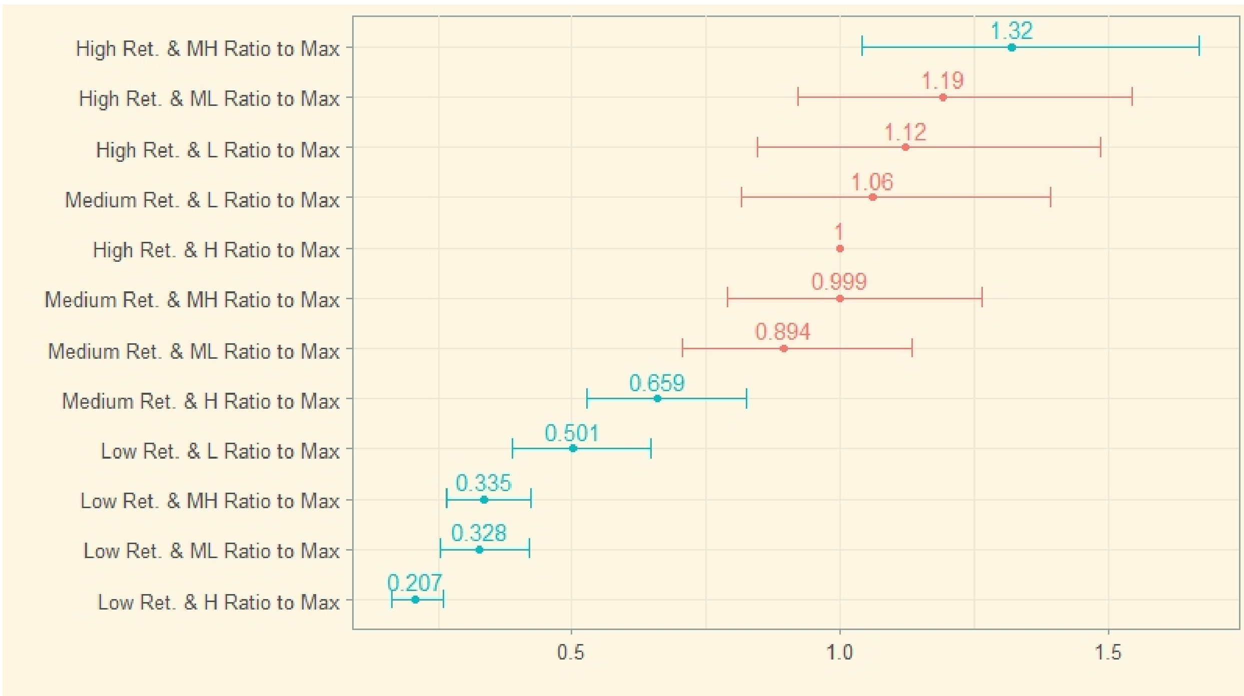

Figure 6: **Odds Ratio Proportional Hazard model: RatioMax and Return.** Odds Ratio from a Proportional Hazard model of the hazard of selling a stock. Each Odds Ratio corresponds to the effect of being in a given category. The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Clustered robust s.e. at bank account level. Blue odds ratios are significantly different from 1 ($p < 0.05$), red are not. Return is the ratio of daily closing price to the purchase price in the investment episode. On the selling date it is equal to the ratio of selling price to the purchase price in the episode. Low [1; 1.06]; Medium (1.06; 1.17]; High (1.17, 5.53]. Ratio to Max Price is the ratio of daily closing price to maximum price up to that time in the investment episode. On the selling date it is equal to the ratio of selling price to maximum price in the episode. Low [0.349; 0.918]; Medium-Low (0.918; 0.957]; Medium-High (0.957; 0.981]; High (0.981; 1]. The baseline category is “High Return and High RatioMax”.

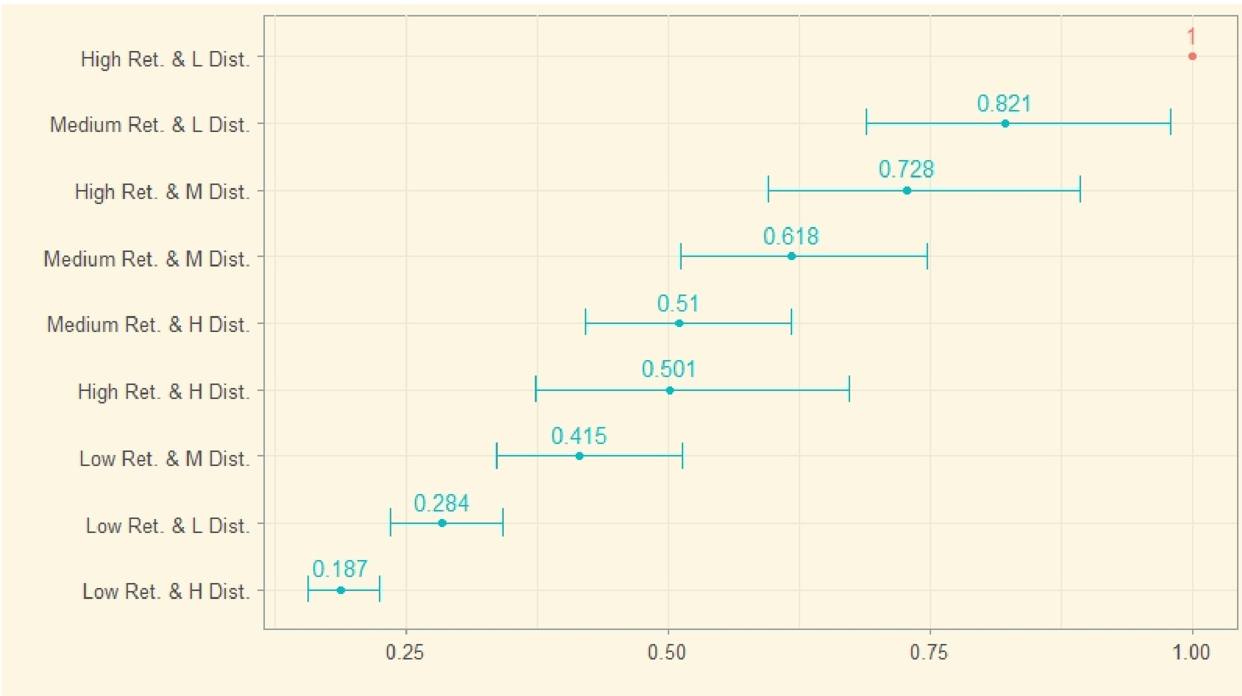

Figure 7: **Odds Ratio Proportional Hazard model: RatioMax and Distance.** Odds Ratio from a Proportional Hazard model of the hazard of selling a stock. Each Odds Ratio corresponds to the effect of being in a given category. The event is the sale. Each day when the stock is not sold is a non-event. Each observation is at stock-bank account-day level. Clustered robust s.e. at bank account level. Blue odds ratios are significantly different from 1 ($p < 0.05$), red are not. Return is the ratio of daily closing price to the purchase price in the investment episode. On the selling date it is equal to the ratio of selling price to the purchase price in the episode. Low [1; 1.06]; Medium (1.06; 1.17]; High (1.17, 5.53]. Distance in Time from Max Day is the rescaled distance as defined in Definition 1. Low [0; 0.07]; Medium [0.07; 0.34); High [0.34; 1]. The baseline category is "High Return and Low Distance".

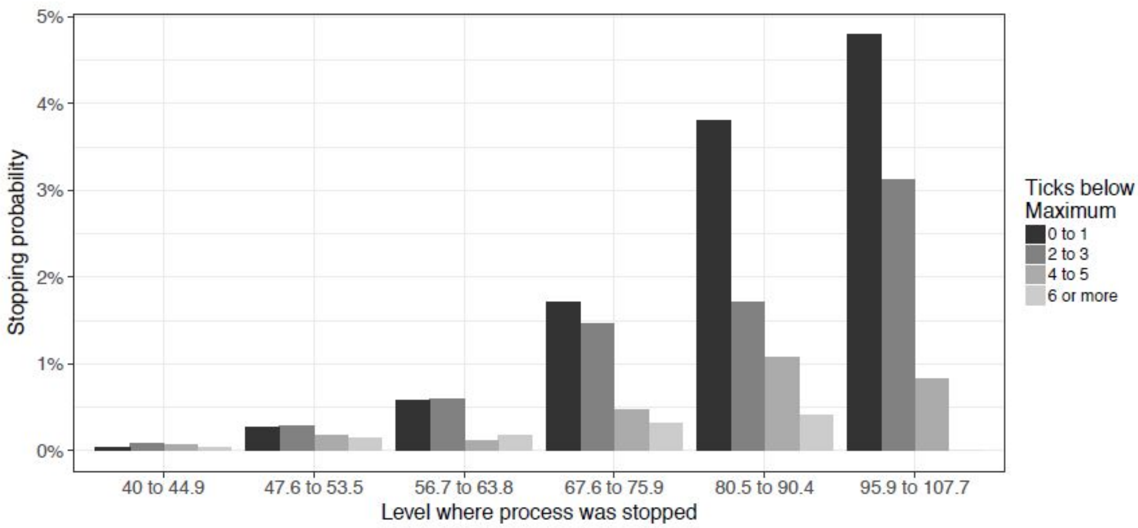

Figure 8: **Experimental Results from Strack and Viefers (2019).** Empirical stopping frequency per any level of the price and distance from past maximum (Figure 4 in Strack and Viefers, 2019).