Reconsideration about a "Jump" in Selling Schedule on Return Since Purchase in Sell-Day Portfolios: An Extension of Ben-David and Hirshleifer (2012)

June 1, 2020

1 Background

Some previous studies on the disposition effect use a sell-day portfolio sample and typically find a "jump" at zero return in the selling schedule as a function of return since purchase (e.g., Kaustia, 2010). However, using daily portfolio sample, Ben-David and Hirshleifer (2012) show that there is no jump in the selling schedule at zero return, and instead, the selling schedule is asymmetric V-shaped for stocks with a short holding period and gets flatter as the holding period increases. That is, given a relatively short holding period, the selling probability increases as an absolute return increases while the sensitivity of the selling probability to the absolute return is grater in the gain domain than in the loss domain. (See Table A1 in Appendix A for review of literature relating to the shape of selling schedule.)

Ben-David and Hirshleifer (2012, p2520) state that "Even if the absence of any sign realization preference, such jumps can easily occur when aggregating over wide ranges of negative and positive profits.", and suggest possible factors making the selling probability to jump at zero return since purchase.

- Restricting a sample to stocks included in sell-day portfolios.
- Averaging selling probabilities across stocks with a different holding period, weighted by the distribution of return since purchase which has a high peak around zero return.
- Using coarse bins.

While Ben-David and Hirshleifer (2012) provide many hints, they do not show detailed mechanism transforming the selling schedule from the asymmetric V-shape in daily portfolios to the step-like shape in sell-day portfolios. This document tries to identify the mechanism, introducing a few additional factors involving.

2   Illustration and Notations

What we will try to show in this document is illustrated in Figure 1. (The values plotted are manually created for illustration purpose and are not from the actual data.[1])

In the figure, both Curves 1 and 2 are inverse V-shaped while the downward curve from the peak at zero return is steeper in the loss domain than in the gain domain in Curve 1 and is steeper in the gain domain than in the loss domain in Curve 2. In addition, in Curves 1 and 2, the downward curve from the peak at zero return is convex both in the gain and the loss domains. Curve 3 is generated by Curve 1 divided by Curve 2. As illustrated, an asymmetric inverse V-shaped curve (Curve 1; the left panel) divided by an oppositely asymmetric inverse V-shapes curve with convex downward curves from the peak at zero (Curve 2; the middle panel) leads to the step-like curve (Curve 3; the right panel).

Note that Curves 1 and 2 do not need to be exactly inverse V-shaped in order for a division of the two curves leading to a step-like shape. As we see below, some conditions being met is sufficient. (The asymmetric inverse V-shape is one example to meet the conditions and also is seen in our empirical data used in the following analysis.)
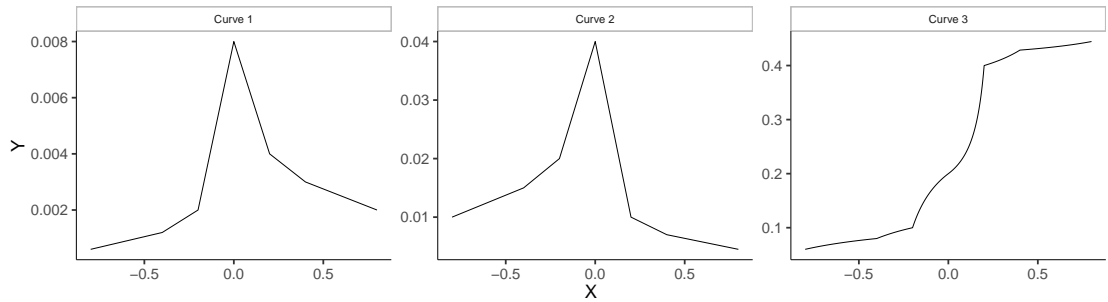

Figure 1: Illustration: An asymmetric inverse V-shape divided by an oppositely asymmetric inverse V-shape leading to a step-like shape.

Table 2 lists notations used in the following discussion.

Table 1: Notations

Variable	Description
N, NG, NL	Number of stocks, gains, and losses in a portfolio, respectively.
Stock i	A stock of interest.
	We are interested in the selling schedule of stock i as a function of the return since purchase of stock i.
Stock j (j =1,2,...N-1)	Stocks other than stock i in the portfolio that stock i belongs to.
P_i^{daily}	Daily probability of stock i being sold.
P_j^{daily}	Daily probability of stock j being sold.
$P_i^{sell\text{-}day}$	Probability of stock i being sold on sell-days.

---

[1]In Figure 1, X={-0.8, -0.4, -0.2, 0, 0.2, 0.4, 0.8}. Y={0.0006, 0.0012, 0.002, 0.008, 0.004, 0.003, 0.002} for Curve 1 and Y={0.01, 0.015, 0.02, 0.04, 0.01, 0.007, 0.0045} for Curve 2.

We will show that a step-like selling schedule on sell-days ($P_i^{sell\text{-}day}$) is led by a division of two asymmetric inverse V-shaped curves (with convex downward curves both in the gain and the loss domains). To do so, in Section 3.2, we first show that P_i^{daily} as a function of return since purchase of stock i is asymmetric inverse V-shaped, and then, in Section 3.3, we show that the probability of stock i being included in a sell-day portfolio as a function of the return since purchase of stock i is oppositely asymmetric inverse V-shaped. Finally, Section 4 shows the mechanism and conditions where a division of two asymmetric inverse V-shaped curves leads to a step-like curve.

3   Empirical Findings

3.1   Data

In this section, we use stock trading data provided by a large mainstream stockbroker in the UK, which include transaction histories of 182,569 accounts for the period between April 2012 and June 2016.

Because purchase prices of stocks bought before the beginning of the data period are unknown, accounts opened before the start of the data period were excluded. Multiple intra-day trades conducted on the same account on the same stock were aggregated with quantity weighted prices. Then, we reconstructed daily portfolio data in which each position is recorded everyday from a first purchase date to a liquidation date. Positions opened on the day and short positions were excluded. The return since purchase was calculated by using a quantity weighted average purchase price of the stock for a given account and a closing price of the stock as of one day prior to the date. Stocks' closing prices were retrieved from the Datastream.[2] Commissions and dividends were not considered in the calculation of returns since purchase. Positions having missing variables in either the price data or the transaction data were excluded from the daily portfolio sample.

As a consequence, the daily portfolio sample includes 8,573,731 account-days for 20,687 accounts (23,687,688 account-stock-days).

3.2   The Effect of Weighted-Averaging across Stocks with Different Holding Period

First, using a small daily portfolio sample consisting of stocks with a holding period no more than 50 days, we illustrate the effect of averaging selling probabilities across stocks which differ in a holding period.

Figure 2 shows the daily selling schedule on return since purchase, P_i^{daily}, separately for five stock groups which differ in holding periods. The curve in each panel is asymmetric V-shaped (i.e., the selling probability increases as an absolute return increases while the positive sensitivity to the absolute return is greater in the gain domain than in the loss domain). In addition, the curve

---

[2]For stocks' identification in data-merging, we used Stock Exchange Daily Official List (SEDOL) numbers which is used for clearing in the United Kingdom and Ireland.

gets flatter as a holding period increases. These are consistent with Figure 1 in Ben-David and Hirshleifer (2012). The asymmetry of the sensitivity of the selling schedule to the absolute return represents the disposition effect in the daily portfolio sample.

Also, comparing across panels in Figure 2, the selling probability tends to be lower as a holding period increases. That is, the longer the holding period the lower the selling probability.[3] [4] In addition, the reduction of the selling probability on holding period is larger when the holding period is relatively short than when it is long. (Compare, for example, the difference in the selling probability between the leftmost and the second leftmost panels with the difference between the rightmost and the second rightmost panels in Figure 2.)
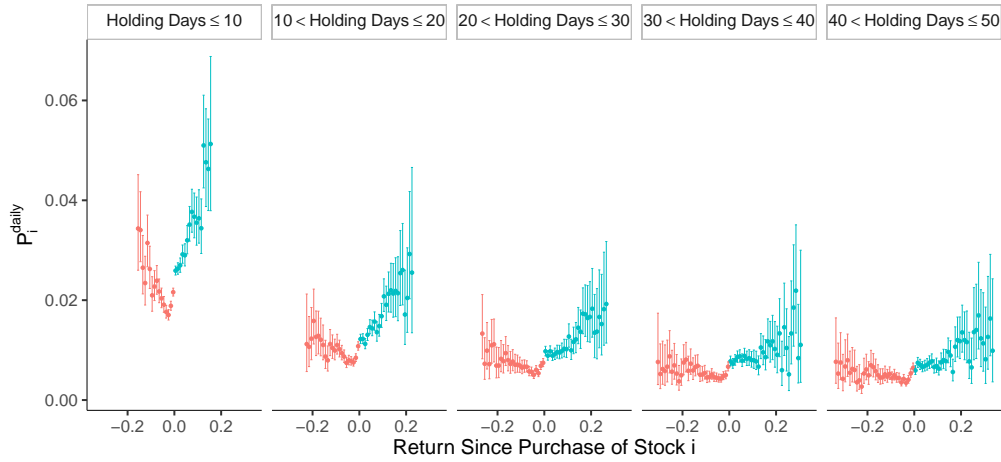

Figure 2: Proportion of stocks being sold for subsets of stocks by holding days (P_i^{daily}; Daily portfolio sample consisting of stock with $Holding\text{-}days \leq 50$). The bin-width is 0.01. The error bars are 95% confidence intervals.

Figure 3 plots the distribution of returns since purchase for each of the five stock groups introduced in Figure 2. The shorter the holding period the more the concentration at zero return and the downward curve from the peak at zero return toward either of the positive or the negative domain is mostly convex. These are because of three factors—(i) the larger the absolute return since purchase the larger the selling probability (seen in Figure 2), (ii) the reduction of selling probability is large when the holding period is short (seen in Figure 2), and (iii) mechanically the large the holding period the larger the absolute return.[5] These three factors together make the reduction of the number of unsold stocks slower as the absolute return increases, leading to the convexity. Note that the distribution

---

[3]The lowest point in the selling schedule in Figure 2 is not at zero return but in a region of small losses. This is because the distribution of returns having a highest peak in a region of small losses is used as a weight when averaging selling schedule within stock group. We will discuss the mechanism later. For example, if we use a sample consisting of stocks with a holding period of exactly 1 day, we see the bottom around zero return.

[4]The shape may relate to the Ostrich effect as briefly discussed below.

[5]Figure A1 shows a relationship between the return since purchase and the holding period.

4

of returns since purchase shown in Figure 3 reflects the selling schedule shown in Figure 2.

Figure 3: Distribution of returns since purchase for subsets of stocks by holding days (Daily portfolio sample consisting of stock with $Holding\text{-}days \leq 50$). The bin-width is 0.25%.

Averaging the selling schedules in Figure 2 across stock groups, weighted by the distribution of returns in Figure 3 leads to Figure 4. The selling schedule in Figure 4 has a peak around zero return. This is because two things work together in the weighted-averaging: (i) As seen in the comparison across panels in Figure 2, stocks with a short holding period has a higher selling probability than stocks with a longer period and (ii) As seen in Figure 3, the distribution of returns of the stocks with a short holding period highly concentrates around zero return.
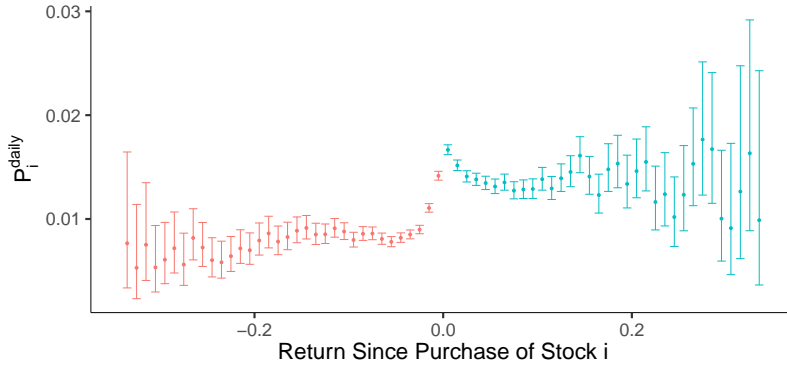

Figure 4: Proportion of stocks being sold (P_i^{daily}; Daily portfolio sample consisting of stock with $Holding\text{-}days \leq 50$). The bin-width is 0.01. The error bars are 95% confidence intervals.

So far, we use the restricted sample including only stocks with a holding period no more than 50 days for illustration purpose. If we do the weighted-averaging

on all daily portfolios[6], we have Figure 5. This is the daily selling schedule on return since purchase for all stocks in the daily portfolio sample.

Similar to Figure 4, the aggregated selling probability seen in Figure 5 has a high peak at zero return and decreases as an absolute return increases toward either of positive or negative side. This is again because (i) the selling probability decreases as a holding period increases and (ii) the distribution of returns becomes less concentrated at zero return as a holding period increases. (i) and (ii) work together in weighted averaging across stocks with a different holding period, resulting in a high peak at zero return. Also, the downward curve from the peak at zero return seen in Figure 5 is convex both in the gain and the loss domains. This is also due to the shape of the distribution of returns since purchase (i.e., the high concentration at zero return and the convex decrease in the selling probability on the absolute return).

In addition, because, as seen in Figure 2, the positive sensitivity of the selling probability to an absolute return is larger in the gain domain than in the loss domain (i.e., V-shape is asymmetric), the downward curve from the peak at zero in Figure 5 is steeper in the loss domain than in the gain domain. (We call this shape "asymmetric inverse V-shape".) That is, while the weighted averaging creates a high peak at zero return in aggregated selling schedule, the negative sensitivity of the selling probability to the absolute return is smaller in the gain domain than in the loss domain because, as seen in Figure 2, the selling probability in each panel increases more in the gain domain than in the loss domain as the absolute return increases.
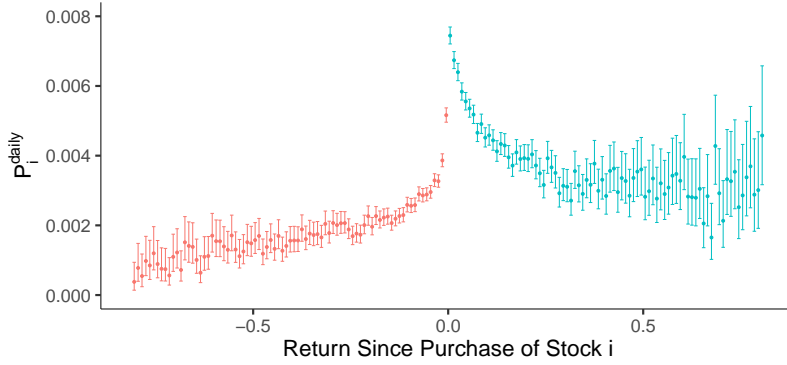

Figure 5: Proportion of stocks being sold (P_i^{daily}; Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

[Interim Summary of Empirical Findings]
As seen in Figure 5, the selling schedule on return since purchase in the daily portfolio sample is asymmetric inverse V-shaped with convex curves both in the gain and the loss domain. This is because the following three factors work together in weighted averaging across stocks with a different holding period.

---

[6]We use portfolios including at least one gain and at least one loss because the sell-day analysis in studies on the disposition effect examines investors' choice among stocks across domains.

(a) The shorter the stock's holding period the higher the daily selling probability of the stock (compare the selling probability across panels in Figure 2).

(b) The distribution of returns since purchase highly concentrates around zero return and becomes less concentrated as a holding period increases (Figure 3).

(c) Given a relatively short holding period, the larger the absolute return the higher the daily selling probability while the sensitivity of the selling probability to an absolute return is larger in the gain domain than in the loss domain (asymmetric V-shape in each panel of Figure 2).

The factors (a) and (b) lead to the inverse V-shape with convex downward curves in Figure 5 and the factor (c) leads to the asymmetry in the sensitivity of the selling schedule to the absolute return (i.e., the asymmetric downward curve from the peak at zero return in Figure 5).[7]

3.3 The Effect of Restricting a Sample to Sell-Day Portfolios

Now we consider the effect of restricting the sample to sell-day portfolios.

Figure 6 shows the proportion of stocks in daily portfolios included in a sell-day portfolio (i.e., the probability of stock i being included in a sell-day portfolio as a function of the return since purchase of stock i). The curve shape is similar to that in Figure 5 (i.e., inverse V-shaped with convex downward curves from the peak). However, as opposed to Figure 5, the downward curve from the peak at zero return is steeper in the gain domain than in the loss domain in Figure 6.
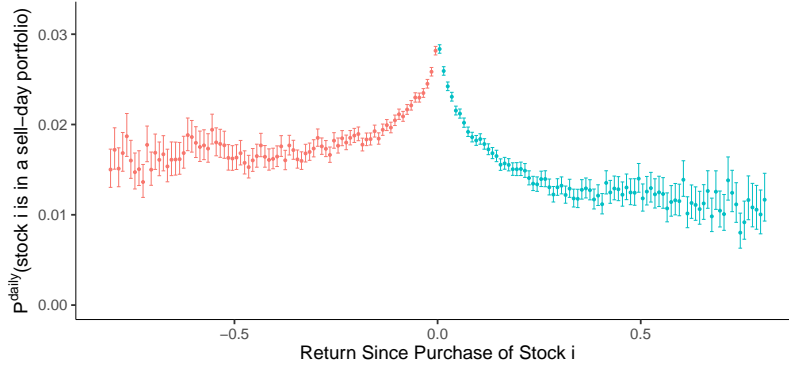

Figure 6: Proportion of stocks included in a sell-day portfolio (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

---

[7]In other words, even with a completely flat daily selling schedule on return since purchase for a given holding period (i.e., even if the curves in Figure 2 were flat), as long as the selling probability decreases as a holding period increases, the weighted-averaging leads to the daily selling schedule on return, averaged across stocks with different holding periods, peaks around zero (i.e., the inverse V-shaped), while the asymmetric V-shaped selling schedule seen in each panel of Figure 2 leads to the asymmetry in Figure 5.

As in many previous studies, if we restrict Figure 5 to sell-day portfolios, we have the step-like shape seen in Figure 7, $P_i^{sell\text{-}day}$. This is a selling schedule on sell-days.

This sample restriction is equivalent to Figure 5, P_i^{daily}, divided by Figure 6, $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$. See Equation 1.

$$P_i^{sell\text{-}day} = \frac{P_i^{daily}}{P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)} \tag{1}$$
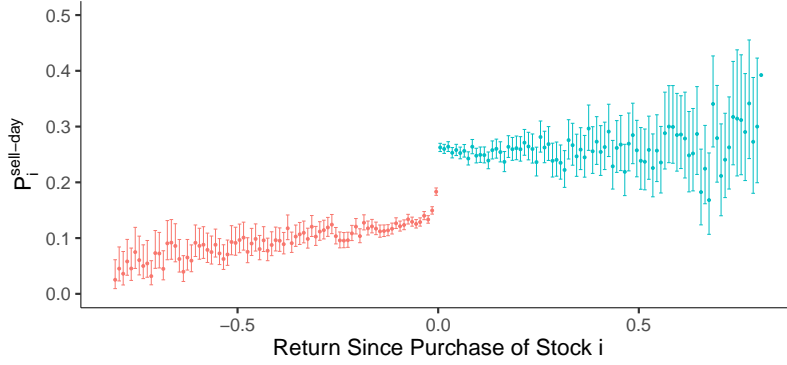

Figure 7: Proportion of stocks being sold ($P_i^{sell\text{-}day}$; Sell-day portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

As discussed above, the asymmetric inverse V-shaped daily probability of stock i being sold (Figure 5) divided by the oppositely asymmetric inverse V-shaped probability of stock i being included in a sell-day portfolio (Figure 6) leads to the step-like shape (Figure 7). (Note that we discuss detailed mechanism about the division of two asymmetric inverse V-shaped curves in Section 4.)

Then, the next questions are:
(1) Why is $P(stock\ i\ in\ a\ sell\text{-}day\ portfolio)$ as a function of return (Figure 6) inverse V-shaped?
(2) Why is the downward curve from the peak at zero return in Figure 6 steeper in the gain domain than in the loss domain?

$P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ can be expressed as Equation 2.

$$P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio) = 1 - (1 - P_i^{daily}) \prod_{j=1}^{N-1} (1 - P_j^{daily}) \tag{2}$$

Equation 2 indicates that $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ increases as

the probability of other stocks in the same portfolio being sold, P_j^{daily}, increases.[8]

If we replace the y-axis of Figure 5, P_i^{daily}, with P_j^{daily}, keeping the x-axis variable unchanged, we have Figure 8. This is the probability of other stocks (stock j) being sold as a function of return of stock i.
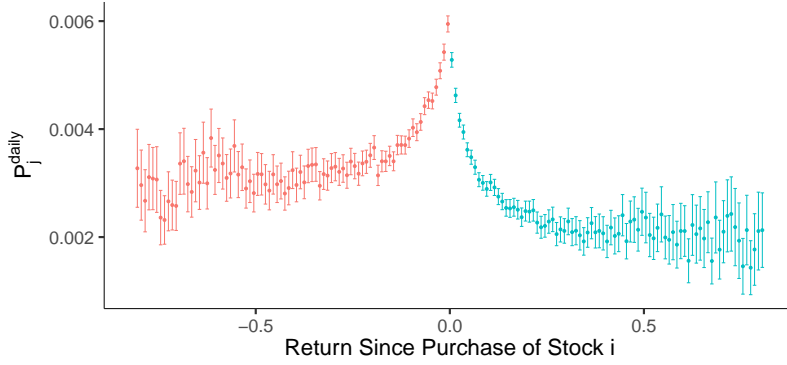

Figure 8: Proportion of other stocks in the portfolios being sold (P_j^{daily}; Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

Similar to Figure 5, Figure 8 is inverse V-shaped. What is a reason of it?

As seen in Figure 9, the holding period of stock j is, on average, similar to that of stock i. Because an absolute return tends to increases as a holding period increases (Figure A1), the larger the absolute return of stock i the larger the absolute return of stock j (Figure 9). Further, because the selling probability is an increasing function of the absolute return (Figure A1), the larger the absolute return of stock i the larger the average selling probability of stocks j, leading to the inverse V-shape in Figure 8.

In short, P_j^{daily} as a function of the return of stock i (Figure 8) has a similar shape to P_i^{daily} (Figure 5) through the similarity in holding periods among stocks in the same portfolio.

---

[8]We assume that all of P_i^{daily} and P_j^{daily} for $j = \{1, 2, ..N\}$ are independent of each other. Also, $P(stock\ i\ in\ a\ sell\text{-}day\ portfolio)$ increases as the N increases.
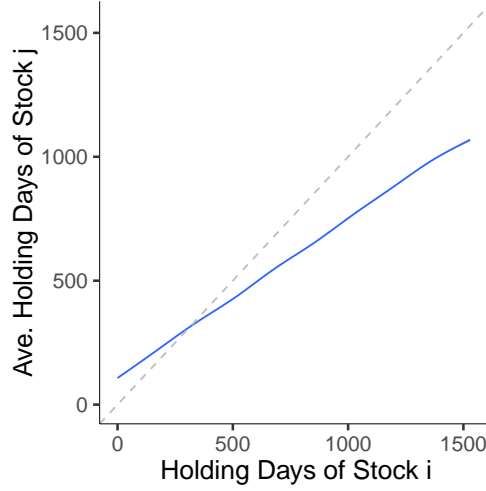
Figure 9: Holding Periods of Stock i and Average Holding Period of Stocks j.

In addition, the downward curve from the peak at zero return in Figure 8 is steeper when the return of stock i is positive than when the return of stock i is negative. This is because, as seen in Figure 10, the sensitivity of an absolute return of stock j to an absolute return of stock i is larger when the return of stock i is negative than when the return of stock i is positive. That is, because the absolute return of stock j increases more as a return of stock i decreases in the loss domain than as a return of stock i increases in the gain domain (the asymmetric V-shape in Figure 10), the selling probability of stock j as a function of return of stock i, seen in Figure 8, deceases less in the loss domain than in the gain domain. (Remind that the larger the absolute return the larger the selling probability.)
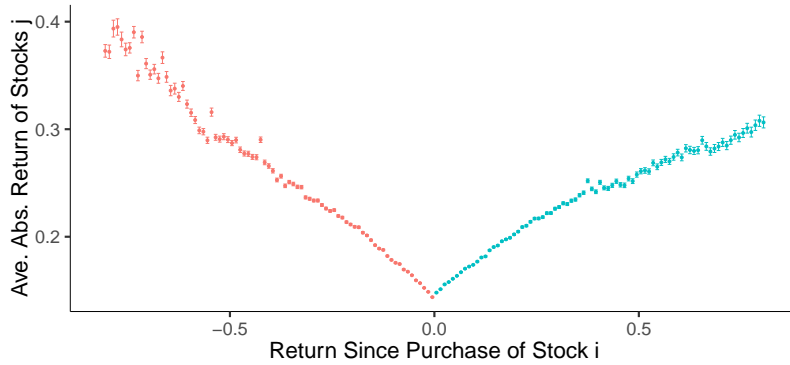

Figure 10: Average absolute return of stocks j as a function of return of stock i (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$).

Why is Figure 10 asymmetric? The reason is quite mechanical.

As the return of stock i decreases, stock i tends to have more stocks j having a return larger than that of stock i. There is no boundary of stock

10

j's return increasing. Conversely, as the return of stock i increases, stock i tends to have more stocks j having a return smaller than that of stock i. However, stock j's return can decrease only up to -1. Therefore, the sensitivity of the average absolute return of stocks j to the absolute return of stock i becomes asymmetric. In short, the curve shape in Figures 8 and 10 is asymmetric because the distribution of returns since purchase is, by nature, positively skewed. (As Equation 2 tells, Figure 6 reflects the curve shape in Figure 8.)

In order to confirm it, we simulate 1,000-days stock prices following the geometric Brownian motion for 10,000 portfolios, each of which consists of five stocks, and plots an average absolute return of stock j on a return of stock i.[9] The result shown in Figure 11 resembles to the curve shape in Figure 10. Even simulated stock prices create the asymmetric V-shape in the average absolute return of stock j as a function of return of stock i seen in Figure 10, indicating that the asymmetry in the average absolute return of stock j as a function of return of stock i is due to mechanical effect (i.e., the positively skewed return distribution).
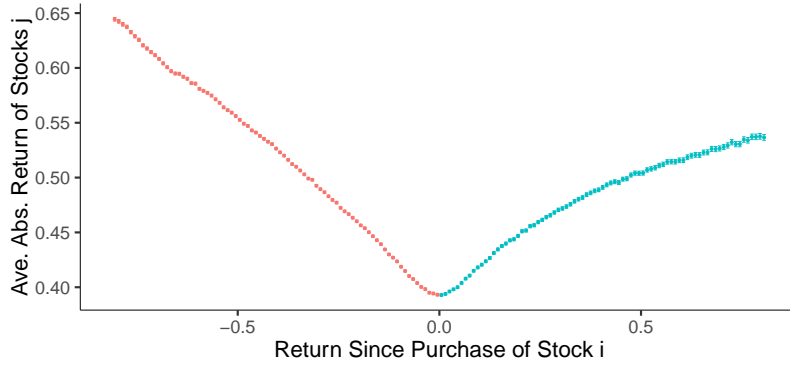

Figure 11: Return of stock i and average absolute return of stocks j (Simulated data).

Furthermore, P(stock i is in a sell-day portfolio) in Equation 2 can be rewriten as in Equation 3.

$$P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio) = P_i^{daily} + (1 - P_i^{daily})B_2, \qquad (3)$$

where B_2 represents the daily probability of at least one of stocks j being sold: $B_2 = 1 - \prod_{j=1}^{N-1}(1 - P_j^{daily})$. See Appendix D for more details.

B_2 as a function of the return since purchase of stock i, by construction, has a similar shape to that of P_j^{daily} seen in Figure 8. As discussed above, this is one reason for the oppositely asymmetric inverse V-shape in $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ seen in Figure 6. In addition, $(1 - P_i^{daily})$ which is multiplied to B_2 in Equation

---

[9]In the same manner as the empirical data, all simulated stocks contribute as stock i once and as stock j four times.

3 plays a role to enhance the asymmetry in B_2, contributing to the asymmetry of $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ which is opposite to the asymmetry of P_i^{daily}.

[Summary of Empirical Findings]
(1) The daily selling schedule on return since purchase, aggregated across stocks, seen in Figure 5 is inverse V-shaped with convex downward curves both in the gain and the loss domain. This is because (a) and (b) below work together in the weighted averaging across stocks with a different holding period.
(a) The shorter the stock's holding period the higher the daily selling probability of the stock. (Compare the selling probability across panels in Figure 2).
and
(b) The distribution of returns since purchase highly concentrates around zero and becomes less concentrated as a holding period increases (Figure 3).

(2) The negative sensitivity of daily selling schedule to an absolute return seen in Figure 5 is larger in the loss domain than in the gain domain because:
(c) Given a relatively short holding period, the larger the absolute return the higher the daily selling probability while the positive sensitivity of the selling probability to an absolute return is larger in the gain domain than in the loss domain (asymmetric V-shape in each panel of Figure 2). (This asymmetry in the sensitivity represents the disposition effect.)

(3) The probability of stock i being included in a sell-day portfolio seen in Figure 6 is inverse V-shaped with convex downward curves from the peak at zero return because:
(d) Stocks in the same portfolio tend to have a similar holding period (Figure 9). That is, due to the similar holding periods, an absolute return of stock i and that of other stocks in the same portfolio (stocks j) are positively correlated, leading to the inverse V-shape in Figure 8. This further leads to the inverse V-shape in Figure 6. (Refer Equation 2.)

(4) The negative sensitivity of the probability of stock i being included in a sell-day portfolio to the absolute return of stock i seen in Figure 6 is larger in the gain domain than in the loss domain because:
(e) The distribution of returns since purchase is, by nature, positively skewed with a negative boundary at -1. (Figures 10 and 11 show the asymmetric sensitivity of an absolute return of stock j as a function of return of stock i.)

(5) The asymmetric inverse V-shaped curve seen in Figure 5 divided by the oppositely asymmetric inverse V-shaped curve seen in Figure 6 leads to the step-like shape in Figure 7. (Note that both Figures 5 and 6 have convex downward curve in the gain and the loss domains.)

Out of five factors listed above, (a) and (c) are behavioral things while (b), (d), and (e) are quite mechanical things.
While "holding period", "absolute return since purchase", and "selling probability"

are highly correlated, the causal direction is unknown. For example, stocks with a long holding period may be initially bought with an intention of a long-term investment, leading to the smaller selling probability. Alternatively, investors may pay less attention to stocks as a holding period increases, leading to the smaller selling probability.

The factors (a) and (c) may relate to investors' attention. That is, just after investors buy stocks (or build a portfolio), they pay much attention to their portfolio with a great sensitivity to price movements. However, as a holding period increases, they pay less attention with a less sensitivity to price movements. In addition, due to the ostrich effect, the sensitivity of attention to price movement is greater for stocks in gain than those in loss.

4   Details of Division of Two Asymmetric Inverse V-shapes

Section 3.2 showed why P_i^{daily} as a function of return since purchase of stock i is asymmetric inverse V-shaped with convex downward curves both in the gain and the loss domain. Section 3.3 showed why $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ as a function of return since purchase of stock i is oppositely asymmetric inverse V-shaped with convex downward curves both in the gain and the loss domain.

In Section 4.1, we review how a division of two lines works, leading to a hyperbola. Then, in Section 4.2, we see the mechanism and conditions where a division of two asymmetric inverse V-shaped probabilities leads to a step-like schedule. Finally, in Section 4.4, we see that P_i^{daily} as a dividend and $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ as a divisor meet those conditions in the empirical data.

4.1   Mathematics of Division of Two Lines

Consider two lines: $y_1 = a_1 + b_1 x$ and $y_2 = a_2 + b_2 x$.

A division of the two lines is:
$y_3 \overset{\text{def}}{=} \frac{y_1}{y_2} = \frac{a_1 + b_1 x}{a_2 + b_2 x} = \frac{b_1}{b_2} + \frac{a_1 - \frac{a_2 b_1}{b_2}}{a_2 + b_2 x}$

This means that y_3 is a rectangular hyperbola with asymptotes of $x = -\frac{a_2}{b_2}$ (vertical line) and $y = \frac{b_1}{b_2}$ (horizontal line).

Let (x_c, y_c) denote the center of the rectangular hyperbola: $(x_c, y_c) \overset{\text{def}}{=} (-\frac{a_2}{b_2}, \frac{b_1}{b_2})$.
Also, let $Z = \frac{a_1 - \frac{a_2 b_1}{b_2}}{b_2}$.
Then, $y_3 = \frac{b_1}{b_2} + \frac{Z}{x + \frac{a_2}{b_2}}$.

This means that the sign of Z determines the location of the hyperbola. That is:
The vertices are $(x_c + \sqrt{Z}, y_c + \sqrt{Z})$ and $(x_c - \sqrt{Z}, y_c - \sqrt{Z})$. for $A > 0$ and
The vertices are $(x_c - \sqrt{-Z}, y_c + \sqrt{-Z})$ and $(x_c + \sqrt{-Z}, y_c - \sqrt{-Z})$. for $Z < 0$.[10]

---

[10]The distance between the center and the vertices is $\sqrt{2|Z|}$. The linear eccentricity is

Note that, when $Z < 0$, the rectangular hyperbola, y_3, locates upper-left and lower-right of the center, $(x_c, y_c) = (-\frac{a_2}{b_2}, \frac{b_1}{b_2})$. This means that the hyperbola is upward at any value of x. We can confirm it by taking the first derivative of y_3.

$$y_3'(x) = \frac{-Z}{(x + \frac{a_2}{b_2})^2}, \; where \; Z = \frac{a_1 - \frac{a_2 b_1}{b_2}}{b_2}. \tag{4}$$

It shows that, as long as $Z < 0$, the slope (i.e., the first derivative line) of y_3 is always positive.

4.2   Division of Two Asymmetric Inverse V-shaped Probabilities

Now we assume that an asymmetric inverse V-shaped curve can be approximated as a joint of many lines. (We can consider it as fitting the data points with many piecewise linear regression lines.)

We denote:
> the return since purchase of stock i as x,
> p_i^{daily} as Y_1,
> $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ as Y_2, and
> $P_i^{sell\text{-}day}$ as Y_3.

We divide the whole range of values of x into many sub ranges $x_k, k = \{..., -2, -1, 1, 2, ...\}$ where a negative value of k indicates all values of x in x_k are negative and a positive value of k indicates all values of x in x_k are positive. We assume that Y_1 can be approximated with many lines, $y_{1(k)}$ corresponding to $x_k, k = \{..., -2, -1, 1, 2, ...\}$. Note that $y_{1(-1)}$ represents a fitted line for a range of negative values of x closest to $x = 0$ and $y_{1(1)}$ represents a fitted line for a range of positive values of x closest to $x = 0$. Similarly, we assume that Y_2 can be approximated with many lines, $y_{2(k)}$ corresponding to $x_k, k = \{..., -2, -1, 1, 2, ...\}$. Finally, We denote a division of two lines, $\frac{y_{1(k)}}{y_{2(k)}}$, as $y_{3(k)}$.

An example of approximation of an inverse V-shaped probability with many lines is illustrated in the left and the middle panels of Figure 12. In the example, both Y_1 and Y_2 are approximated by six lines $y_{1(k)}, k = \{-3, -2, -1, 1, 2, 3\}$ and $y_{2(k)}, k = \{-3, -2, -1, 1, 2, 3\}$. The right panel is a result of dividing Y_1 by Y_2.

$2\sqrt{|Z|}$ and the eccentricity is $\sqrt{2}$. The radius of curvature at the vertices is $\sqrt{2|Z|}$. Refer https://en.wikipedia.org/wiki/Hyperbola.
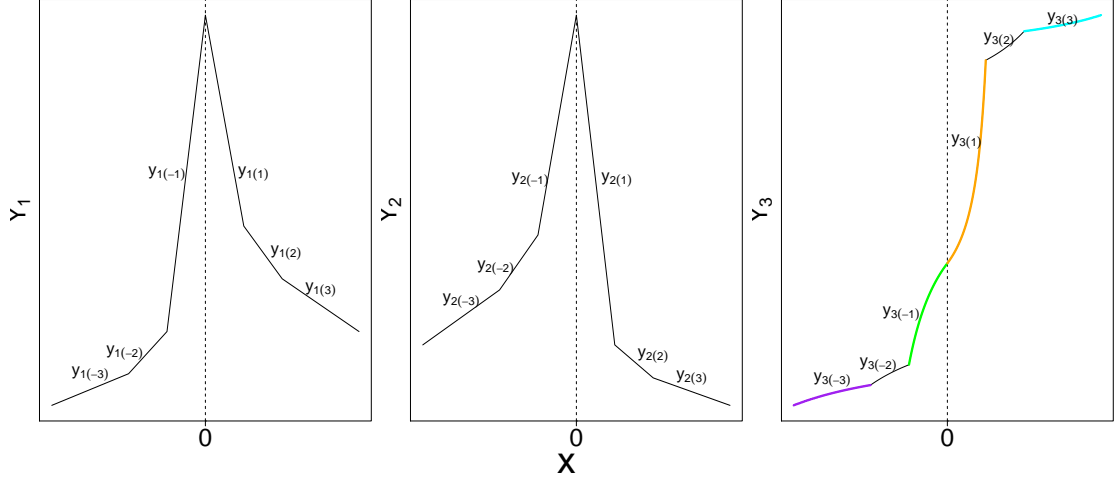
Figure 12: Illustration: An asymmetric inverse V-shape divided by an oppositely asymmetric inverse V-shape leading to a step-like shape. The colored parts in the right panel corresponds to the same-colored parts in the bottom panels of Figure 13.

In addition, Figure 13 illustrates how the division of two lines leads to a part of a hyperbola. In each column, the blue line divided by the red line in the top panel results in the colored part of the rectangular hyperbola in the bottom panel. The colored part in the bottom panels corresponds to the same-colored curve in the right panel of Figure 12. Note that, because Y_1 and Y_2 are probabilities, $a_{1(k)} > 0$ and $a_{2(k)} > 0$. Also, the inverse V-shape for both Y_1 and Y_2 means that $b_{1(k)} > 0$ and $b_{2(k)} > 0$ for $x \leq 0$, and $b_{1(k)} < 0$ and $b_{2(k)} < 0$ for $x > 0$.
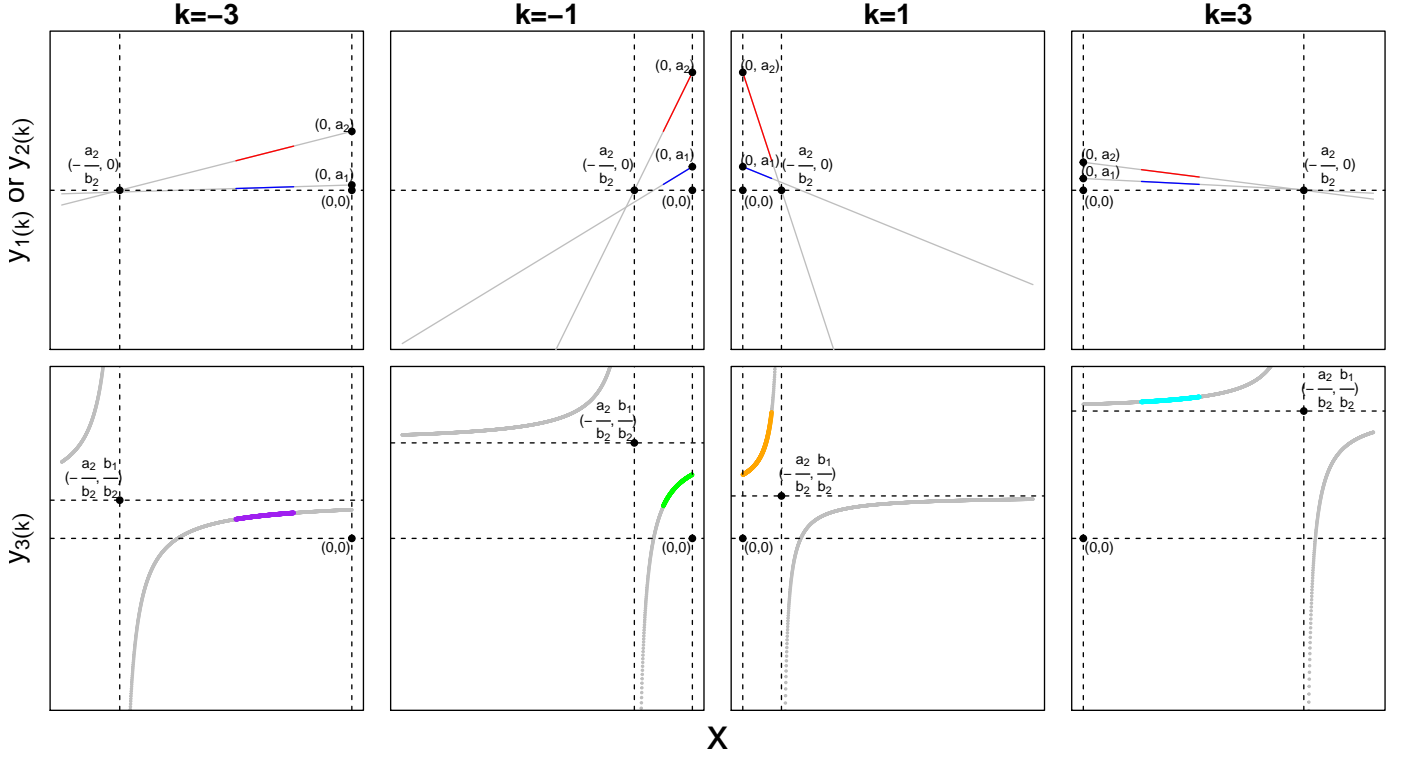
Figure 13: Illustration: Division of two asymmetric inverse V-shaped curve approximated as a joint of many lines. The blue line divided by the red line in the top panels leads to the colored parts of the curve in the bottom panels. The colored parts in the bottom panels corresponds to the same-colored parts in the right panel of Figure 12. In the inside-figure texts, the (k) notations are excluded for simplicity.

4.3 Conditions for Division of Two Inverse V-Shaped Probability leading to a Step-Like Shape

First, we consider a condition for $y_{3(k)}$ being upward.
As seen in Equation 4, when $Z < 0$, $y_{3(k)}$ is upward. Equation 4 can be expanded as:

$$Z = \frac{a_{1(k)} - \frac{a_{2(k)}b_{1(k)}}{b_{2(k)}}}{b_{2(k)}} = \frac{\frac{a_{1(k)}}{b_{1(k)}} - \frac{a_{2(k)}}{b_{2(k)}}}{\frac{b_{2(k)}}{b_{1(k)}}} \tag{5}$$

Because $b_{1(k)} > 0$ and $b_{2(k)} > 0$ for $x \leq 0$, and $b_{1(k)} < 0$ and $b_{2(k)} < 0$ for $x > 0$, the denominator of Equation 5 is always positive. Thus, the condition for $Z < 0$ is $-\frac{b_{2(k)}}{a_{2(k)}} < -\frac{b_{1(k)}}{a_{1(k)}}$ (Condition 1). Note that $-\frac{b_{1(k)}}{a_{1(k)}}$ and $-\frac{b_{2(h)}}{a_{2(h)}}$ are the x-intercept of $y_{1(k)}$ and $y_{2(h)}$, respectively.

Condition 1: $-\frac{b_{2(k)}}{a_{2(k)}} < -\frac{b_{1(k)}}{a_{1(k)}}$ for $y_{3(k)}$ being upward.

16

Geometrically, Condition 1 means that the x-intercept of $y_{2(k)}$ is more far from $x = 0$ than the x-intercept of $y_{1(k)}$ for $x \leq 0$, and conversely, the x-intercept of $y_{1(k)}$ is more far from $x = 0$ than the x-intercept of $y_{2(k)}$ for $x > 0$.

Further, when Condition 1 is met, the rectangular hyperbola, $y_{3(k)}$, locates upper-left and lower-right of the center, $(x_c, y_c) = (-\frac{a_{2(k)}}{b_{2(k)}}, \frac{b_{1(k)}}{b_{2(k)}})$, and thus, is upward anywhere. In addition, because $y_{2(k)}$ is a probability, it should not be negative. Therefore, in the loss domain, x should be greater than the x-intercept of $y_{2(k)}$, $-\frac{a_2}{b_2}$, which is the vertical asymptote of the rectangular hyperbola. Thus when Condition 1 is met for $x < 0$, the corresponding branch of $y_{3(k)}$ locates lower-right of the center (the fourth quadrant).[11] Similarly, in the gain domain, x should be less than the x-intercept of y_2, $-\frac{a_2}{b_2}$, which is vertical asymptote of the rectangular hyperbola. Thus when Condition 1 is met for $x > 0$, the corresponding branch of $y_{3(k)}$ locates upper-left of the center (the second quadrant). Refer Figure 13.

Next, we consider what determines the steepness of $y_{3(k)}$. The denominator Equation 4, $x + \frac{a_{2(k)}}{b_{2(k)}}$, indicates that $y_{3(k)}$ gets steeper as the distance between x and the x-intercept of $y_{2(k)}$ decreases. This reflects the fact that a hyperbola, $y_{3(k)}$, gets steeper as x gets closer to the vertical asymptote.

The numerator of Equation 4 is expanded to Equation 5. The numerator of 5 indicates that Z gets more negative as the distance of x-intercepts of two lines, $-\frac{a_{2(k)}}{b_{2(k)}}$ and $-\frac{a_{1(k)}}{b_{1(k)}}$ gets larger. Therefore, for $x \leq 0$, the curve of Y_3 tends to be flatter as $|x|$ increases when Y_1 is more convex than Y_2. Conversely, for $x > 0$, the curve of Y_3 tends to be flatter as $|x|$ increases when Y_2 is more convex than Y_1.

At an intersection of two hyperbolas, $x = x* \leq 0$, as long as Y_2 is convex, $-\frac{a_{2(k-1)}}{b_{2(k)}} > -\frac{a_{2(k-1)}}{b_{2(k)}}$. Thus, the denominator of 4 gets larger as $|x|$ increases, leading to the slope of $y_{3(k-1)}$ tending to be flatter than that of $y_{3(k)}$.

Branches of two rectangular hyperbolas which have a vertical asymptote and locate in the same quadrant have at most two intersections in the quadrant. It can be geometrically said that, at an intersection (or one closer to the asymptotes in the case of two intersections) of such branches of two rectangular hyperbolas, one having a larger horizontal distance between its vertical asymptote and the intersection should be flatter than the other. Thus, the slope of Y_3 gets flatter when x passes each kink point from the origin as long as Y_2 is convex.

Note that, within each rectangular hyperbola y_3, the curve gets steeper as $|x|$ increases. This is due to the shape of rectangular hyperbola with negative Z. However, given that the range of x for each y_3 is small, the change in the slope within y_3 is visually not large (see the right panel of Figure 12).

Condition 2: .

---

[11]The x-intercept of the rectangular hyperbola is $x = -\frac{a_{1(k)}}{b_{1(k)}}$.

If Condition 1 is met for each of y_3, Y_3 is upward. If Condition 2 is met, Y_3 is getting flatter as $|x|$ increase. Therefore, if both conditions are met, Y_3 is upward around $x = 0$ and becomes flatter as x decreases in the loss domain or increases in the gain domain. In the case that a degree of the flattening of Y_3 is large, Y_3 becomes close to a step-like shape.

Seeing it conversely, Y_3 tends to be steeply upward around zero if the fitted line, y_2, closest to the origin in either side is steep.

4.4 Connecting to the Empirical Findings

In this section, we examine whether conditions shown in the previous section are met in the empirical data.

4.4.1 Visual Inspection

Condition 1 can be re-write as $\frac{b_{2(k)}}{a_{2(k)}} < \frac{b_{1(k)}}{a_{1(k)}}$.

Condition 1 means that, when separately plotting the two lines, y_1 and y_2, with a free y-scale ranging from 0 to a_1 or a_2, the slope of the line is visually steeper for y_1 than for y_2 in the loss domain and is visually negatively steeper for y_2 than for y_1 in the gain domain.

According to a visual inspection, Condition 1 seems to be met in the loss domain, particularly in the region of small losses. (Compare Figure 5 (as $y_1 = a_1 + b_1 x$) with the corresponding part of Figure 6 (as $y_2 = a_2 + b_2 x$).) This leads to the upward curve in the region of small losses in Figure 7 (as y_3). However, the difference in steepness between y_1 and y_2 in the region of small gains is not so obvious comparing with the region of small losses. This leads to the flatter curve in the region of small gains in Figure 7.

Also, 6 (as Y_2) are convex in the loss and the gain domains, and thus, Condition 2 seems to be met. This leads to the flat curve in the region of large absolute returns in Figure 7 (as Y_3).

4.4.2 Inspection Based on Piecewise Regression Lines

We conduct a piecewise linear regression on the binned data shown in Figures 5 (as y_1) and 6 (as y_2).[12] The regression is conducted separately for the loss and the gain domains of each figure (i.e., four regressions).

The resulting fitted lines are shown in the left and the middle panels in Figure 14. The right panel plots data points and lines calculated by the left panel data divided by the middle panel data. Table 2 shows whether the fitted lines meet the conditions for ranges of x segmented at each kink points in y_1 and y_2. Condition 1 is met for most values of x, except four short ranges of x. In addition, both b_1 and b_2 get smaller in the absolute term as $|x|$ increases, indicating that Condition 2 is met. Table 2 also shows values of y_3'. It is obvious

---

[12]The algorithm shown by Muggeo et al. (2008) was used. The number of kinks was set at 2 (i.e., 3 lines).

that the y_3' for $x = [-0.019, 0.000]$. This is reflected to the steep upward curve in the region of small losses in the right panel of Figure 14.
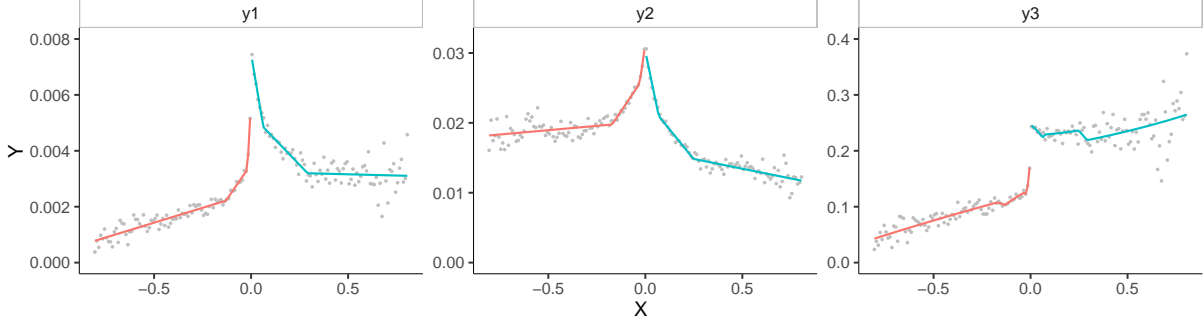

Figure 14: Fitted lines from piecewise linear regressions. In the left and the middle panels, the dots are taken from the binned data which are identical to those in Figure 5 for y_1 and are identical to those in Figure 6 for y_2 while the solid lines are fitted lines based on piecewise linear regressions. In the right panel, the dots and the lines are calculated by y_1 divided by y_2 at the corresponding value of x in the left and the middle panels, respectively.

Table 2: Results of Piecewise Linear Regressions

Range of x	a_1	b_1	a_2	b_2	Condition 1	Range of y_3'
[-0.800, -0.171]	0.002	0.002	0.020	0.002	Y	[0.094, 0.111]
[-0.171, -0.130]	0.002	0.002	0.027	0.041		[-0.118, -0.100]
[-0.130, -0.030]	0.004	0.010	0.027	0.041	Y	[0.190, 0.270]
[-0.030, -0.019]	0.004	0.010	0.031	0.193		[-0.556, -0.478]
[-0.019, 0.000]	0.006	0.130	0.031	0.193	Y	[3.003, 3.858]
[0.000, 0.063]	0.007	-0.041	0.030	-0.134		[-0.529, -0.274]
[0.063, 0.069]	0.005	-0.007	0.030	-0.134	Y	[1.042, 1.120]
[0.069, 0.248]	0.005	-0.007	0.023	-0.034	Y	[0.029, 0.058]
[0.248, 0.291]	0.005	-0.007	0.016	-0.006		[-0.411, -0.398]
[0.291, 0.800]	0.003	0.000	0.016	-0.006	Y	[0.072, 0.110]

5 Demonstration by Simulation

5.1 Method for the Simulation

Finally, we replicate the empirical findings with simulated data where stock prices follow the geometric Brownian motion and investors' selling decisions are determined by the return since purchase and the number of days since purchase, together with XXX model free parameters.

In the simulation, each day of 5 years, xxx investors open a trading account. They construct an initial portfolio consisting of 5 stocks (i.e., $252 \times 5 = 1,260$ days) within 60-days investment window since the account open date. They buy one stock on a day which is randomly chosen within the investment window. The

stock prices follows the geometric Brownian motion with the drift parameter (μ) and the volatility parameter (σ). We set $\mu = 0$ and draw σ from an approximated distribution of annualized standard deviations (i.e., volatility) of daily returns taken from the empirical data, shown in Figure **??**.[13]

Investors' selling decisions are determined by the model introduced below. We assume no additional purchase after the initial portfolio is constructed. Then, we extract investor-days where at least one gain and one loss are in the portfolio, and compute the selling schedule on return in a daily portfolio sample and that in a sell-day portfolio sample.

5.2   Model of Selling Decisions

We introduce a model of selling decisions where the selling probability on each investor-account-day is determined by the return since purchase of stock i ($Return_i$) and the number of days since purchase of stock i ($Days_i$). The main features we observed in the empirical data are:
(1) The selling probability increases as an absolute return increases while the sensitivity of the selling probability is larger for the gain than for losses (i.e., an asymmetric V-shape). (2) The selling probability decreases as a holding period increases while the sensitivity of the selling probability is larger in short holding periods than in long holding periods. In addition, the decay of the probability is lager for losses than for gains.

In order to reflect the feature (1), our model employs the logistic function shown in Equation 6. In the equation, α represents the selling probability when $Return_i = 0$. β is a sensitivity parameter. $0 \leq \gamma \leq 1$ when $Return_i \leq 0$, otherwise $\gamma = 1$. Thus, the sensitivity is attenuated for losses.

$$f(Return_i) = \alpha + \frac{1}{1 + |Return_i|^{-\gamma\beta}} - 0.5 \tag{6}$$

In order to reflect the feature (2), our model employs the double exponential decay function shown in Equation 7. In the equation, π is a minimal selling probability for gains after many days passed since purchase (i.e., $\theta = 1$) and $\theta\pi$ is that for losses ($0 \leq \theta \leq 1$). ϕ represents the proportion of total decay which the fast decay attributes to. S_1 and S_2 are parameters measuring a stability of retention for the fast decay and for the slow decay, respectively ($S_1 < S_2$).[14]

$$d(Days_i) = \pi + \theta\phi(1 - \pi)e^{\frac{-(Days_i - 1)}{S_1}} + (1 - \theta\phi)(1 - \pi)e^{\frac{-(Days_i - 1)}{S_2}} \tag{7}$$

The two functions together determine the selling probability as a function of $Return_i$ and $Days_i$ with xxx free parameters.

$$P_i^{Daily} = f(Return_i)d(Days_i) \tag{8}$$

---

[13]We assume zero correlation in price movement across stocks.
[14]Refer https://en.wikipedia.org/wiki/Forgetting_curve.

In the model, two free parameters, γ and θ, lead to the disposition effect. The parameter, γ, measures an attenuation of investors' sensitivity to the return since purchase when the stock is in loss while the parameter, θ, measures an attenuation of investors' minimum attention paid to the stock when the stock is in loss.

In the simulation, we set parameter values as follows:
$\alpha =$, $\beta =$, $\gamma =$, $\theta =$, $\pi =$, ϕ, $S_1 =$, $S_2 =$.
HS: In the current version, the parameter values were arbitrarily chosen just to have a reasonable results. Ideally, we should conduct an optimization on empirical data to get the parameter values. However, the optimization does not work well. Therefore the results are not model fitting but just a demonstration for creating a step-like selling schedule by simulation.

Figure **??** shows the shape of the double decay function and Figure **??** shows the shape of the logistic function. Figure **??** shows the selling schedule determined by these two functions.

5.3   Results from the Simulation

6   Discussion

HS: We had better discuss why the model assumptions, in particular $0 \leq \gamma \leq 1$ and $0 \leq \theta \leq 1$, are psychologically plausible (if possible).

Appendix A   Literature Review by Ben-David and Hirshleifer (2012)

Ben-David and Hirshleifer (2012) reviewed previous studies about a shape of the selling probability as a function of return since purchase. Table A1 summarizes Section 5 of Ben-David and Hirshleifer (2012).

Table A1: Studies on a Shape of Selling Schedule on Return Since Purchase

| Shape	Study	Reference	Data	Sample/Method
Jump at Zero	Kaustia (2010)	Figure 2	Finnish	Logistic regression on sell-days on each subset by holding periods. (A jump is predicted only for stocks with a short holding period.)
	Grinblatt and Keloharju (2001)	Figure 1 (A)	Finnish	Histogram of the number of sold stocks on return bins. (The histogram is conditional on the stocks being sold).
V-Shape	Ben-David and Hirshleifer (2012)	Whole paper	US	Daily analysis by holding periods (discontinuity analysis, etc).
	Barber and Odean (2013)	Figure 1	US, Finnish	Hazard ratio for sale of stocks (Cox regression; thus the analysis is conditional on stocks not being sold in prior holding period.) The curve is asymmetric V-shape in the US data (Panel A) while the curve is steeply upward for small gains in Finnish data (Panel B).
	Seru et al. (2009)	Figure 3	Finnish	Hazard ratio for sale of stocks (Cox regression; thus the analysis is conditional on stocks not being sold in prior holding period.) (The curve is nearly flat in the loss domain and upward in the gain domain without a jump at zero)
	Grinblatt et al. (2012)	—	Finnish	Hirshleifer mentioned this paper but V-shape is shown only in pre-publication version which we cannot see.
Inverse V-Shape	Odean (1998)	Table VII	US	PGR and PLR on sell-days

Appendix B   Supplemental Figures
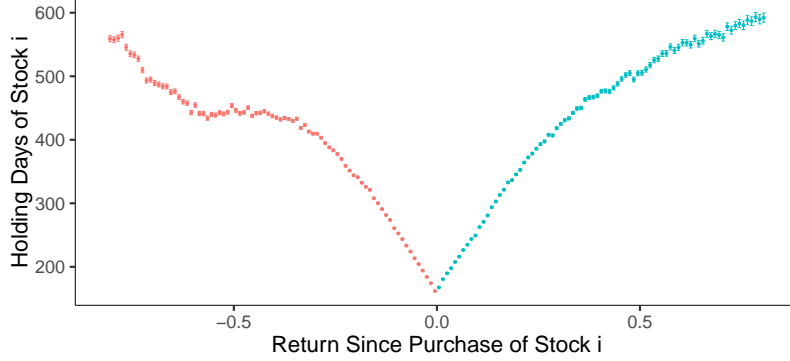
Figure A1: Holding Days of Stock i (Daily portfolio sample). The bin-width is 0.01. The error bars are 95% confidence intervals.

Appendix C   Return Distribution

The distribution of returns since purchase shown in Figure 3 already reflects the selling schedule. To illustrate it, we simulate stock prices for 10,000 stocks for 5 years following the geometric Brownian motion (Annualized $\mu = 0$ and $\sigma = 0.4$). The left panel of Figure A2 shows the distribution of daily stock returns of those 10,000 stocks for 5 years. By the nature of the Brownian motion, the daily returns are normally distributed. The middle panel shows the distribution of returns since purchase assuming that 10,000 investors bought one of the 10,000 stocks and never sold it. The right panel shows the distribution of returns since purchase assuming that 1% of investors sold (liquidated) a stock each day.

The shape of the distribution (i.e., the extent of the concentration around zero return and the convexity of each of the gain and the loss domain) highly depends on the selling probability.
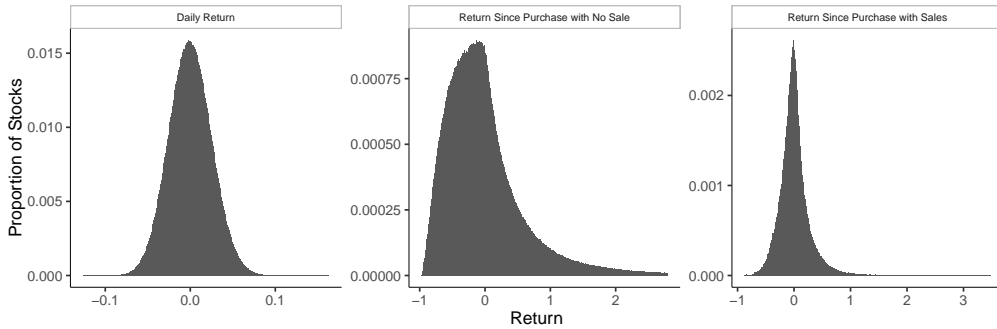

Figure A2: The left, the meddle, and the right panels show the distribution of daily returns, returns since purchase with no sale, and returns since purchase with sales, respectively. The bin-width is 0.001, The extreme 2.5% returns were excluded from the middle and the right panels.

Appendix D More Details of Conditioning on Sell-Days

Further, $P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio)$ can be broken into three parts:
$P(stock\ i\ is\ in\ a\ sell\text{-}day\ portfolio) = A + B + C$.

(A) Probability of only stock i is sold on the day:

$$A = P_i^{daily} \prod_{j=1}^{N-1} (1 - P_j^{daily}) \tag{A1}$$

(B) Probability of at least one of stocks j is sold and stock i is not sold on the day.

$$B = (1 - P_i^{daily})[1 - \prod_{j=1}^{N-1} (1 - P_j^{daily})] \tag{A2}$$

(C) Probability of stock i is sold and at least one of stocks j is sold and stock i is not sold on the day.

$$C = P_i^{daily}[1 - \prod_{j=1}^{N-1} (1 - P_j^{daily})] \tag{A3}$$

Also, B can be divided into two parts: $B = B_1 B_2$.

(B_1) Probability of stock i is not sold.

$$B_1 = 1 - P_i^{daily} \tag{A4}$$

(B_2) Probability of at least one of stocks j is sold.

$$B_2 = 1 - \prod_{j=1}^{N-1} (1 - P_j^{daily}) \tag{A5}$$

Because $P_i^{daily} = A + C$, all the above together,

$$P_i^{sell\text{-}day} = \frac{A + C}{A + B + C} = \frac{P_i^{daily}}{P_i^{daily} + B} = \frac{P_i^{daily}}{P_i^{daily} + (1 - P_i^{daily})B_2} \tag{A6}$$
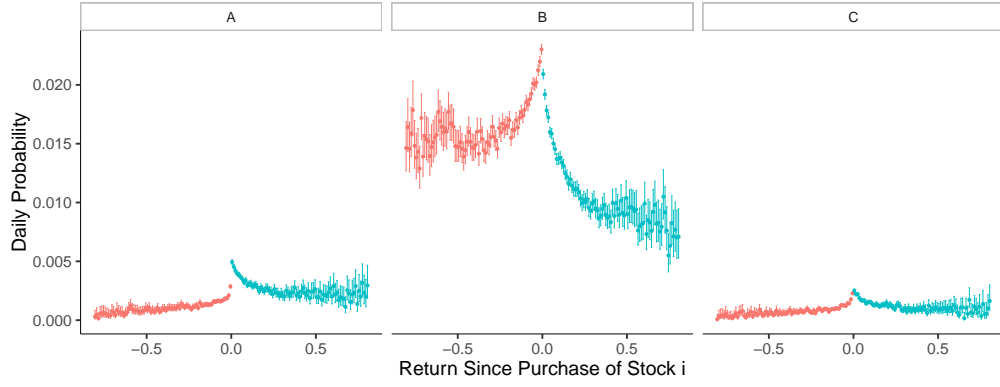
Figure A3: A, B, and C (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.
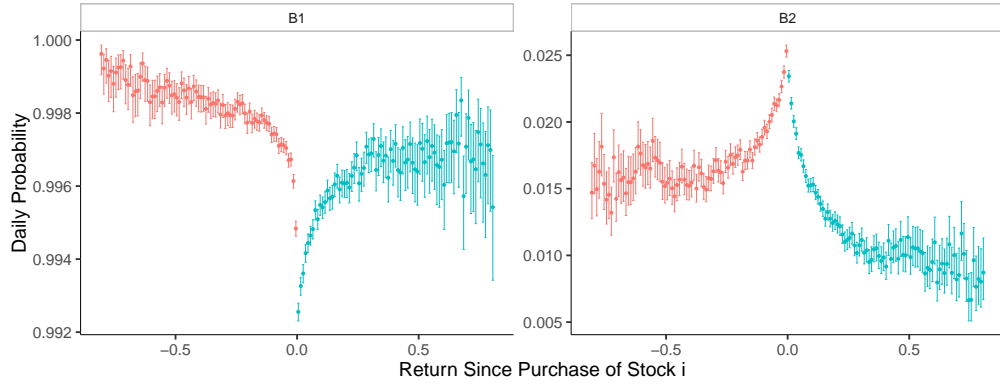

Figure A4: B1 and B2 (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals. $B_1 = 1 - P_i^{daily}$.

References

Brad M Barber and Terrance Odean. The behavior of individual investors. In *Handbook of the Economics of Finance*, volume 2, pages 1533–1570. Elsevier, 2013.

Itzhak Ben-David and David Hirshleifer. 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, 2012.

Mark Grinblatt and Matti Keloharju. What makes investors trade? *J. Financ.*, 56(2):589–616, 2001.

Mark Grinblatt, Matti Keloharju, and Juhani T Linnainmaa. Iq, trading behavior, and performance. *Journal of Financial Economics*, 104(2):339–362, 2012.

M. Kaustia. Prospect theory and the disposition effect. *J. Financ. Quant. Anal.*, 45(3):791–812, 2010.

Vito MR Muggeo et al. Segmented: an r package to fit regression models with broken-line relationships. *R news*, 8(1):20–25, 2008.

T. Odean. Are investors reluctant to realize their losses? *J. Financ.*, 53(5): 1775–1798, 1998.

Amit Seru, Tyler Shumway, and Noah Stoffman. Learning by trading. *The Review of Financial Studies*, 23(2):705–739, 2009.