

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

May 4, 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 greater 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.)

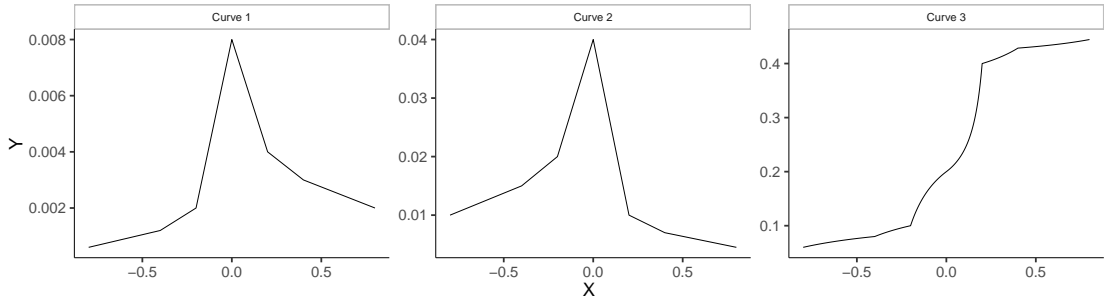


Figure 1: Illustration: An asymmetric inverse V-shape divided by an oppositely asymmetric inverse V-shape leading to a step-like shape.

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 ($j = 1, 2, \dots, N-1$)	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.
$P_i^{sell-day}$	Probability of stock i being sold on sell-days.

¹In 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 ($P_i^{sell-day}$) 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

²As a stock identification for 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.)

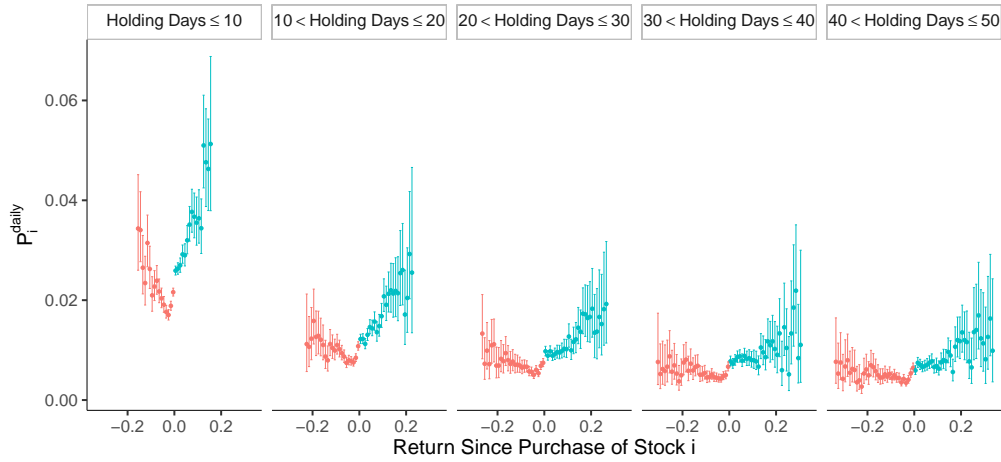


Figure 2: Proportion of stocks being sold for subsets of stocks by holding days (P_i^{daily} ; Daily portfolio sample consisting of stock with $Holding-days \leq 50$). The bin-width is 0.01. The error bars are 95% confidence intervals.

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 larger 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

³The 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.

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

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

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 $Holding-days \leq 50$). 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 holding 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.

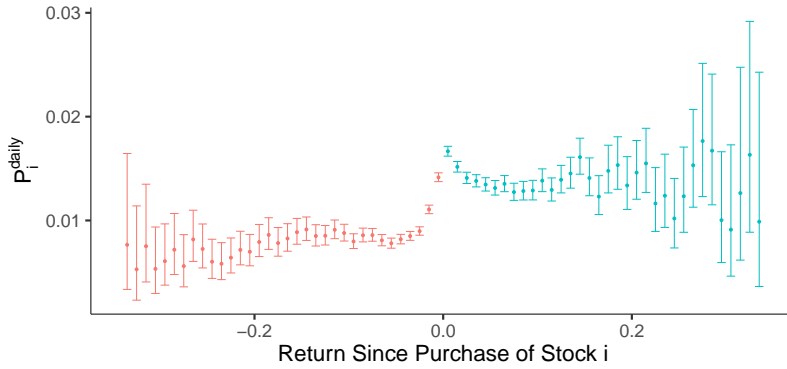


Figure 4: Proportion of stocks being sold (P_i^{daily} ; Daily portfolio sample consisting of stock with $Holding-days \leq 50$). The bin-width is 0.01. The error bars are 95% confidence intervals.

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 portfolios⁶, 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.

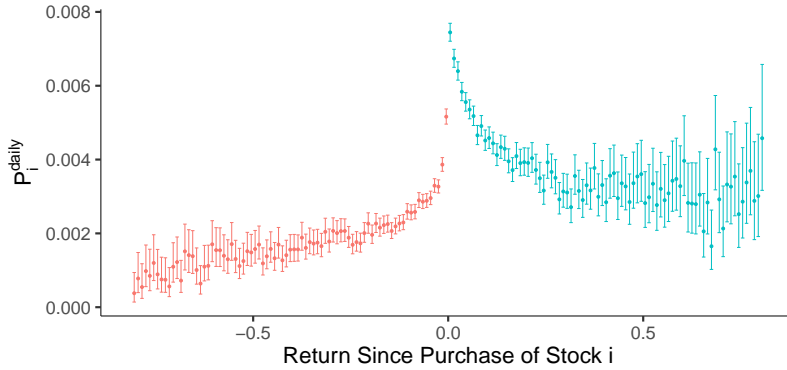


Figure 5: Proportion of stocks being sold (P_i^{daily} ; Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

[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.

⁶We 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.

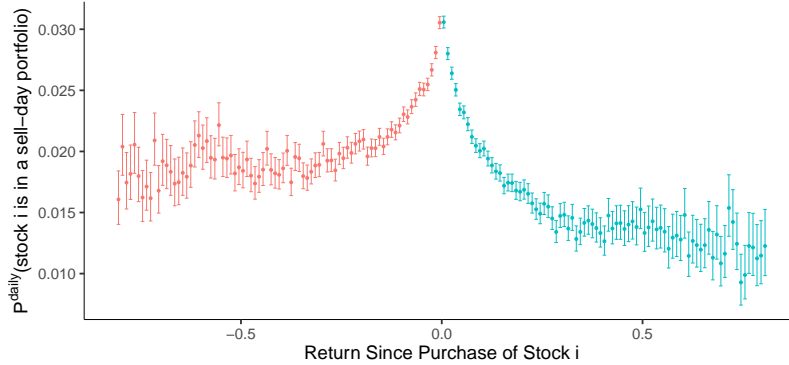


Figure 6: Proportion of stocks included in a sell-day portfolio (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

⁷In 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, $P_i^{sell-day}$. This is a selling schedule on sell-days.

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

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

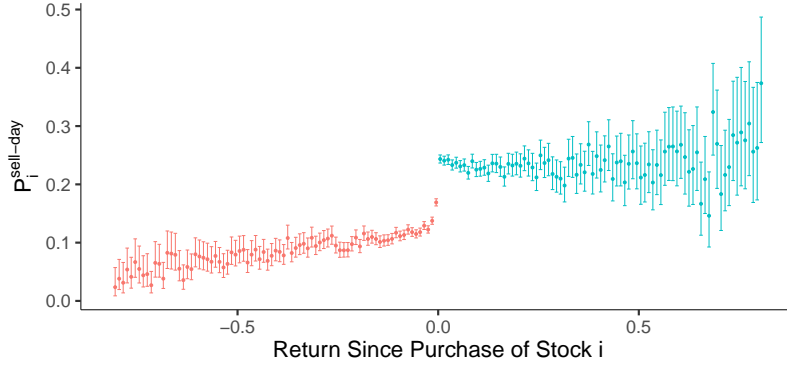


Figure 7: Proportion of stocks being sold ($P_i^{sell-day}$; Sell-day portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

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 $P(\text{stock } i \text{ in a sell-day portfolio})$ 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?

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

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

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

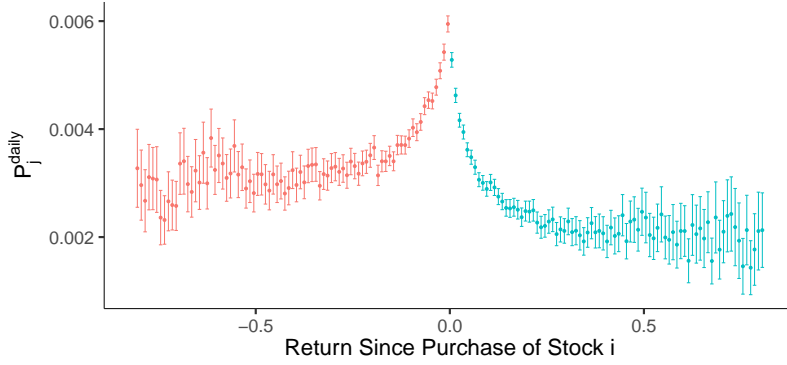


Figure 8: Proportion of other stocks in the portfolios being sold (P_j^{daily} ; Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

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 increase 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.

⁸We assume that all of P_i^{daily} and P_j^{daily} for $j = \{1, 2, \dots, N\}$ are independent of each other. Also, $P(\text{stock } i \text{ in a sell-day portfolio})$ increases as the N increases.

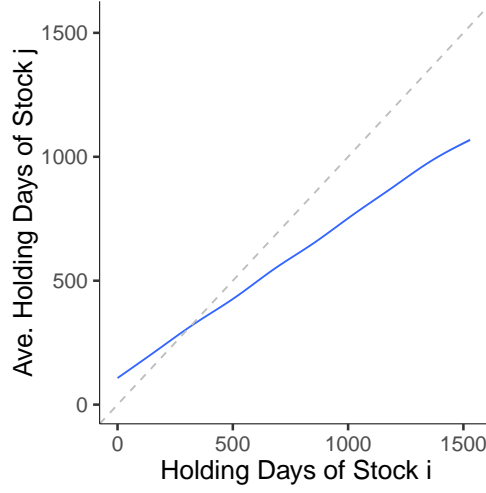


Figure 9: Holding Periods of Stock i and Average Holding Period of Stocks j .

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, decreases less in the loss domain than in the gain domain. (Remind that the larger the absolute return the larger the selling probability.)

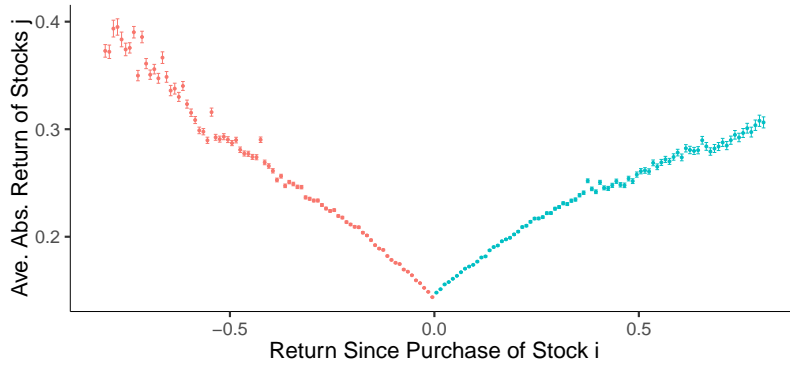


Figure 10: average absolute return of stocks j as a function of return of stock i (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$).

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

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).

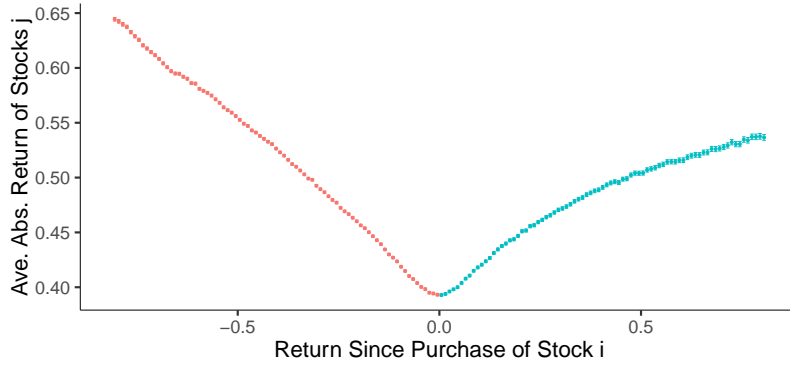


Figure 11: Return of stock i and average absolute return of stocks j (Simulated data).

Furthermore, $P(\text{stock } i \text{ is in a sell-day portfolio})$ in Equation 2 can be rewritten as in Equation 3.

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

where B_2 represents the daily probability of at least one of stocks j being sold: $B_2 = 1 - \prod_{j=1}^{N-1} (1 - P_j^{\text{daily}})$. 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 $P(\text{stock } i \text{ is in a sell-day portfolio})$ seen in Figure 6. In addition, $(1 - P_i^{\text{daily}})$ which is multiplied to B_2 in Equation

⁹In 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 $P(\text{stock } i \text{ is in a sell-day portfolio})$ 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. Section 3.3 showed why $P(\text{stock } i \text{ is in a sell-day portfolio})$ as a function of return since purchase of stock i is oppositely asymmetric inverse V-shaped (with convex downward curves from a peak at zero return to either of the gain or 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 4.3, we see that P_i^{daily} as a dividend and $P(\text{stock } i \text{ is in a sell-day portfolio})$ as a divisor meet those conditions.

4.1 Mathematics of Division of Two Lines

Consider two lines: $y_1 = a_1 + b_1x$ and $y_2 = a_2 + b_2x$.

A division of the two lines is: $y_3 \stackrel{\text{def}}{=} \frac{y_1}{y_2} = \frac{a_1+b_1x}{a_2+b_2x}$.

In the long division format it is: $b_2x + a_2 \overline{)b_1x + a_1}$

Proceeding a first step in the long division leads to:

$$y_3 \stackrel{\text{def}}{=} \frac{a_1+b_1x}{a_2+b_2x} = \frac{b_1}{b_2} + \frac{a_1 - \frac{a_2b_1}{b_2}}{a_2+b_2x}$$

($\frac{b_1}{b_2}$ and $a_1 - \frac{a_2b_1}{b_2}$ can be considered as a quotient and a remainder, respectively)

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.

$$(x_c, y_c) \stackrel{\text{def}}{=} \left(-\frac{a_2}{b_2}, \frac{b_1}{b_2}\right).$$

Also, let $A = \frac{a_1 - \frac{a_2b_1}{b_2}}{b_2}$.

Then, $y_3 = \frac{b_1}{b_2} + \frac{A}{\frac{a_2}{b_2} + x}$

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

4.2 Division of Two Asymmetric Inverse V-shaped Probabilities

Now we approximate an asymmetric inverse V-shaped curves as a joint of many lines. Continuing the above notations and consider that:

x is a small range of return since purchase of stock i ,

y_1 is p_i^{daily} for the range of x ,

y_2 is $P(\text{stock } i \text{ is in a sell-day portfolio})$ for the range of x , and

y_3 is $P_i^{sell-day}$ for the range of x .

Note that, by construction, y_2 is always larger than y_1 (i.e., $y_2 > y_1$ for $N \geq 2$). This leads to $a_2 > a_1$.

In this illustrative example, as seen in the left and the middle panels in Figure 12, we approximate the upward curve in the negative domain with three lines and the downward curve in the positive domain with three lines. In each domain, the line is steeper around $x = 0$ and gets flatter as $|x|$ increases (i.e., convex).

In the negative domain, $b_1 > 0$ and $b_2 > 0$.

Therefore, $A = \frac{a_1 - \frac{a_2 b_1}{b_2}}{b_2}$ is negative only when $\frac{b_2}{a_2} < \frac{b_1}{a_1}$ (Condition 1).

Condition 1 means that, when separately plotting y_1 and y_2 for $x < 0$ 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 .

When Condition 1 is met, $A < 0$. Then, the rectangular hyperbola, y_3 , locates upper-left and lower-right of the center, $(x_c, y_c) = (-\frac{a_2}{b_2}, \frac{b_1}{b_2})$.

In addition, because $y_2 \geq 0$ and $b_2 > 0$, x should be greater than the vertical asymptote, $-\frac{a_2}{b_2}$. (I.e., Because y_2 is a probability, it should not be negative.) Thus, when Condition 1 is met, corresponding y_3 locates lower-right of the center.

Conversely, in the positive domain, $b_1 < 0$ and $b_2 < 0$.

Therefore, $A = \frac{a_1 - \frac{a_2 b_1}{b_2}}{b_2}$ is negative only when $\frac{b_2}{a_2} < \frac{b_1}{a_1}$ (Condition 2). Note that the inequality is the same to Condition 1. However, as opposed to Condition 1, the sign of b_1 and b_2 are negative

Condition 2 means that, when separately plotting y_1 and y_2 for $x > 0$ with a free y-scale ranging from 0 to a_1 or a_2 , the slope of the line is visually negatively

¹⁰The distance between the center and the vertices is $\sqrt{2A}$. The linear eccentricity is $2\sqrt{A}$ and the eccentricity is $\sqrt{2}$. The radius of curvature at the vertices is $\sqrt{2A}$. Refer <https://en.wikipedia.org/wiki/Hyperbola>.

steeper for y_2 than for y_1 .

When Condition 2 is met, $A < 0$. Then, the rectangular hyperbola, y_3 , locates upper-left and lower-right of the center, $(x_c, y_c) = (-\frac{a_2}{b_2}, \frac{b_1}{b_2})$.

In addition, because $y_2 \geq 0$ and $b_2 < 0$, x should be less than the vertical asymptote, $-\frac{a_2}{b_2}$. Thus, when Condition 2 is met, corresponding y_3 locates upper-left of the center.

A division of two lines resulting in a part of a rectangular hyperbola is illustrated in Figure 13. In each column, the blue line divided by the red line in the top panel results in the colored part of the 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.

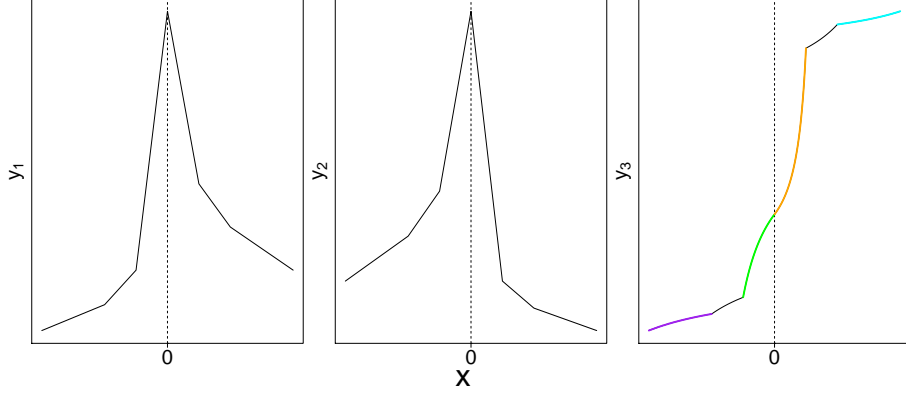


Figure 12: Illustration: An asymmetric inverse V-shape divided by an oppositely asymmetric inverse V-shape leading to a step-like shape. The colored parts in the right panel corresponds to the same-colored parts in the bottom panels of Figure 13.

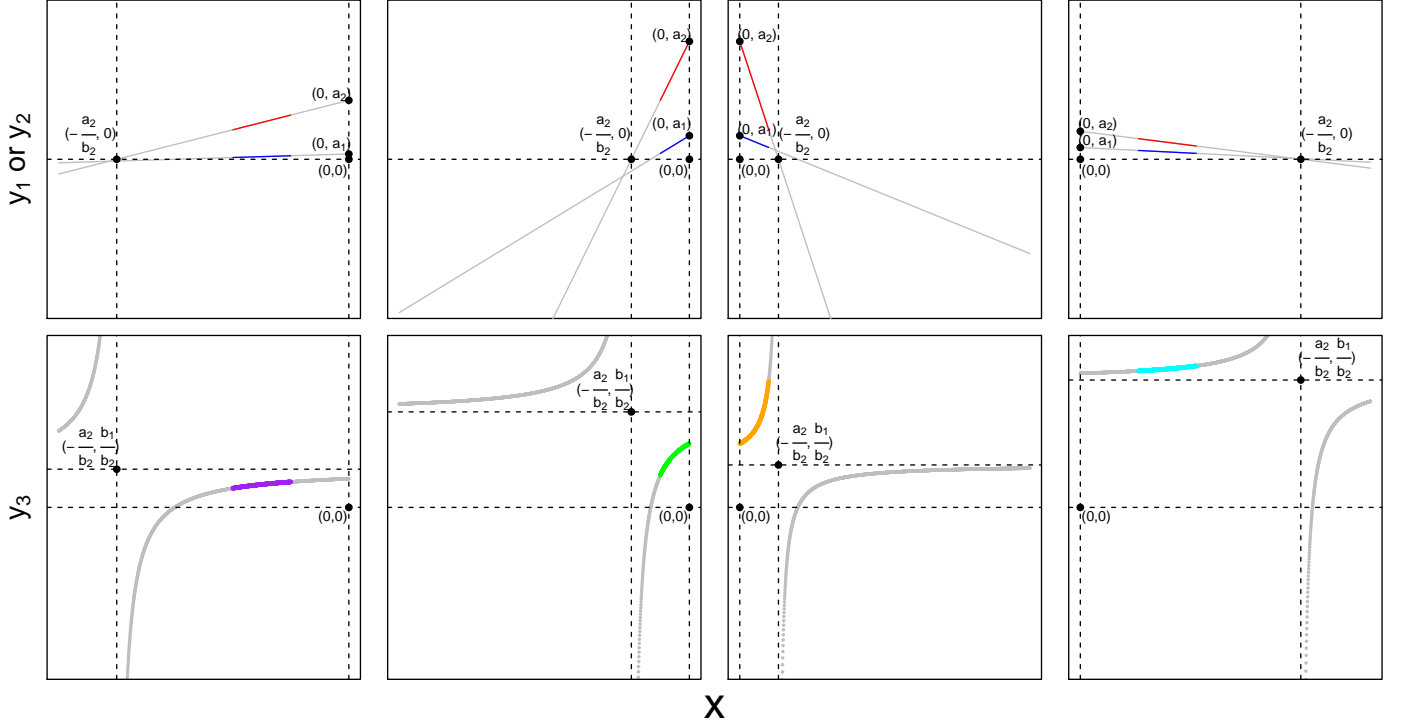


Figure 13: Illustration: Division of two asymmetric inverse V-shaped curve approximated as a joint of many lines. The blue line divided by the red line in the top panels leads to the colored parts of the curve in the bottom panels. The colored parts in the bottom panels corresponds to the same-colored parts in the right panel of Figure 12.

The discussion above shows that the curve of y_3 is upward because it is a part of a rectangular hyperbola with a negative A (thus locating in the first and the fourth quadrant in a coordinate having an origin at the center of the hyperbola). However, in the negative domain of Figure 13, the green curve is much steeper than the purple curve, and similarly, in the positive domain, the orange curve is much steeper than the light-blue curve. What is a reason behind it?

The convex downward curves of y_2 from the peak at $x = 0$ to either of negative or positive side means that $-\frac{a_2}{b_2}$ (i.e., the x -intercept of the line) gets larger in an absolute term as $|x|$ increases. Therefore, in the negative domain, the vertical asymptote, $y = -\frac{a_2}{b_2}$, gets more far from the origin, $(x, y) = (0, 0)$, as x gets more negative, and thus, the corresponding part of the hyperbola is likely to be flatter as x decreases. Similarly, in the positive domain, the vertical asymptote, $y = -\frac{a_2}{b_2}$, gets more far from the origin, $(x, y) = (0, 0)$, as x gets more positive, and thus, the corresponding part of the hyperbola is likely to be flatter as x increases. Therefore, an additional condition for y_1 divided by y_2 leading to a step-like shape is that the curve of y_2 is convex both in the negative and the positive domain (Condition 3).

In addition note that, as $|A|$ gets smaller (i.e., as $\frac{b_2}{a_2}$ gets closer to $\frac{b_1}{a_1}$), the hyperbola becomes closer to a horizontal line, $y = \frac{b_1}{b_2}$, except the area very close to $-\frac{a_2}{b_2}$. Therefore, when $\frac{b_2}{a_2} \approx \frac{b_1}{a_1}$, y_3 is nearly a horizontal line for most values of x . This is another reason for the flat part of y_3 (i.e., $P_i^{sell-day}$).

4.3 Connecting to the Empirical Findings

As seen above,

Condition 1: $\frac{b_2}{a_2} < \frac{b_1}{a_1}$ for $b_1 > 0$ and $b_2 > 0$.

Condition 2: $\frac{b_2}{a_2} < \frac{b_1}{a_1}$ for $b_1 < 0$ and $b_2 < 0$.

Condition 3: y_2 is convex both in the negative and the positive domain.

Condition 1 means that, when separately plotting the two lines in the negative domain 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 .

Condition 2 means that, when separately plotting the two lines in the positive domain with a free y-scale ranging from 0 to a_1 or a_2 , the slope of the line is visually negatively steeper for y_2 than for y_1 .

According to a visual inspection, Condition 1 seems to be met in an arbitrary range of negative x , particularly in the region of small losses. (Compare the loss domain of Figure 5 (as $y_1 = a_1 + b_1x$) with the corresponding part of Figure 6 (as $y_2 = a_2 + b_2x$.) This leads to the upward curve in the region of small losses in Figure 7.

Similarly, Condition 2 seems to be met in an arbitrary range of positive x (Compare the gain domain of Figure 5 (as $y_1 = a_1 + b_1x$) with the corresponding part of Figure 6 (as $y_2 = a_2 + b_2x$.) However, the difference in steepness between two curves is not so obvious comparing with the loss domain. This leads to a flatter line in the gain domain in Figure 7.

Also, as seen in Figure 6, Condition 3 seems to be met. This leads to a flatter line in the region of large absolute returns in Figure 7.

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	<i>PGR</i> and <i>PLR</i> on sell-days

Appendix B Supplemental Figures

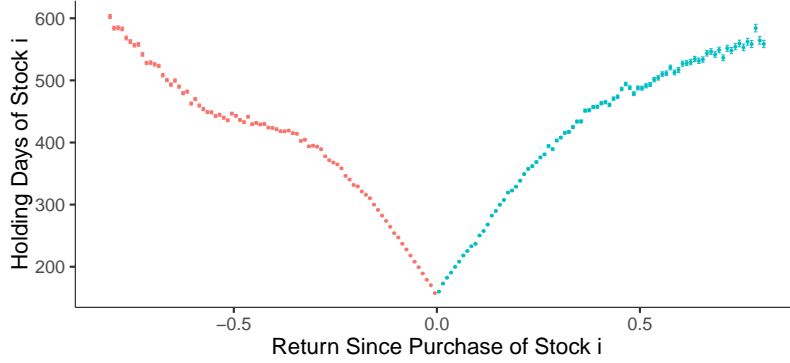


Figure A1: Holding Days of Stock i (Daily portfolio sample). The bin-width is 0.01. The error bars are 95% confidence intervals.

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.

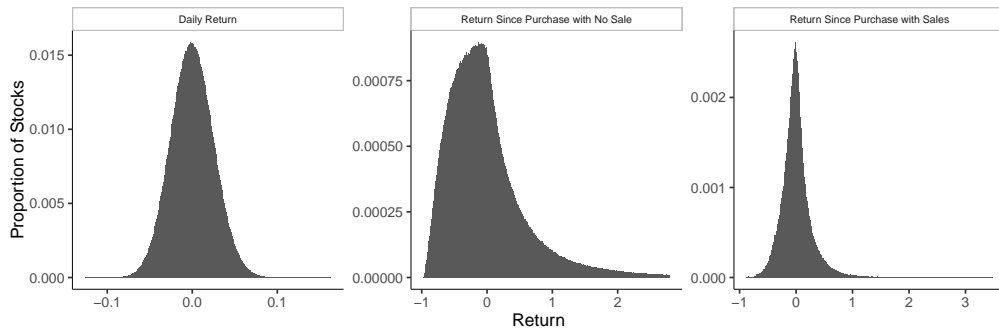


Figure A2: The left, the middle, and the right panels show the distribution of daily returns, returns since purchase with no sale, and returns since purchase with sales, respectively. The bin-width is 0.001. The extreme 2.5% returns were excluded from the middle and the right panels.

Appendix D More Details of Conditioning on Sell-Days

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

$$P(\text{stock } i \text{ is in a sell-day portfolio}) = A + B + C.$$

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

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

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

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

(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.

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

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

(B₁) Probability of stock i is not sold.

$$B_1 = 1 - P_i^{\text{daily}} \quad (\text{A4})$$

(B₂) Probability of at least one of stocks j is sold.

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

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

$$P_i^{\text{sell-day}} = \frac{A + C}{A + B + C} = \frac{P_i^{\text{daily}}}{P_i^{\text{daily}} + B} = \frac{P_i^{\text{daily}}}{P_i^{\text{daily}} + (1 - P_i^{\text{daily}}) B_2} \quad (\text{A6})$$

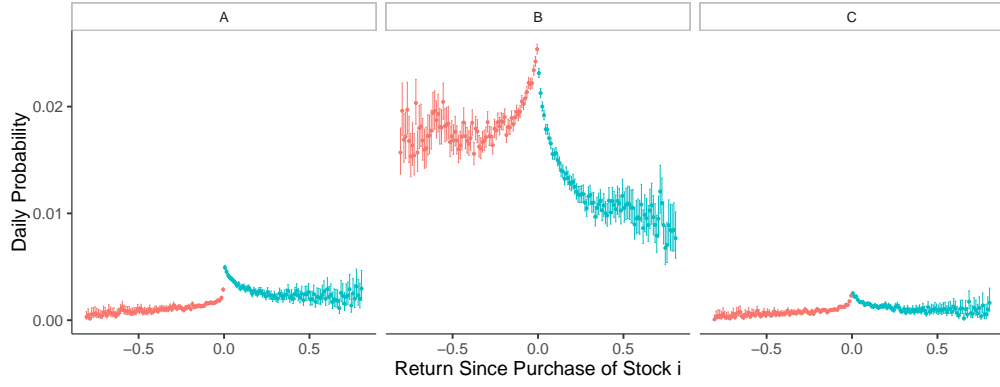


Figure A3: A, B, and C (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals.

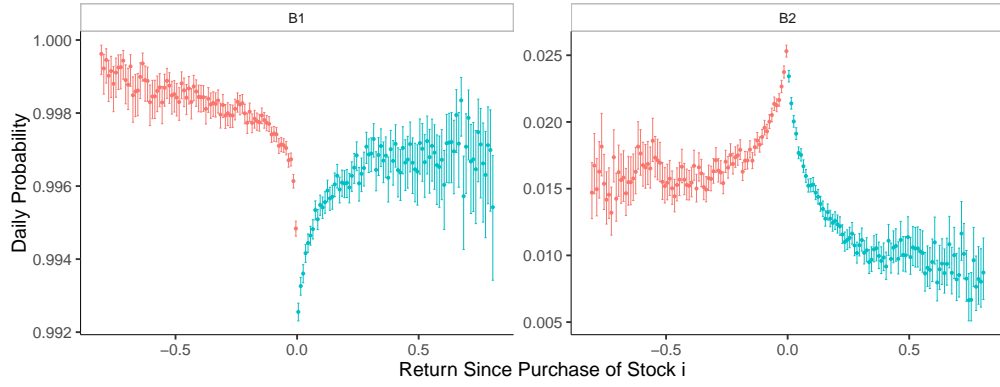


Figure A4: B1 and B2 (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). The bin-width is 0.01. The error bars are 95% confidence intervals. $B_1 = 1 - P_i^{daily}$.

Appendix E Unused Figures

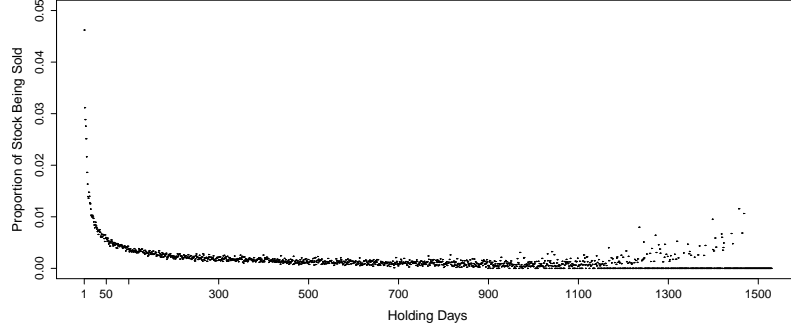


Figure A5: Proportion of stocks being sold (P_i^{daily} ; Daily portfolio sample) as a function of holding days.

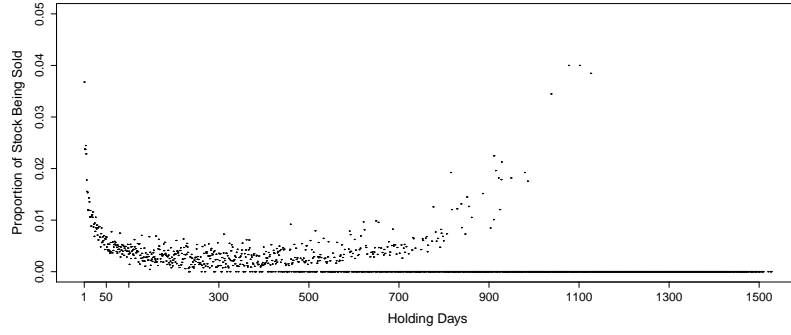


Figure A6: Proportion of stocks being sold (P_i^{daily} ; Daily portfolio sample) as a function of holding days for stocks whose return is between -0.01 and 0.01.

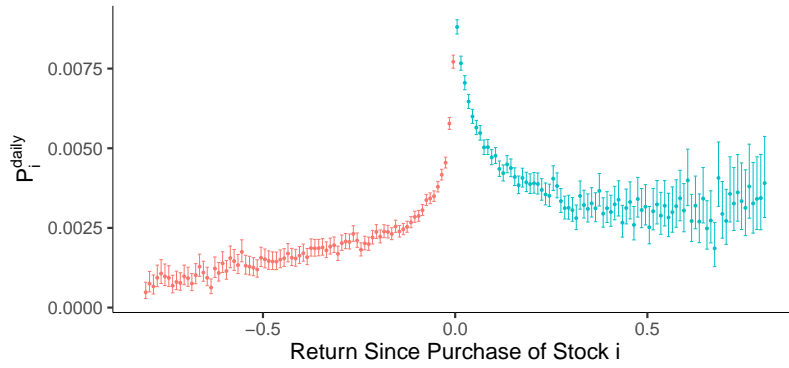


Figure A7: Proportion of stocks being sold (P_i^{daily} ; Daily portfolio sample). The bin-width is 0.01. The error bars are 95% confidence intervals.

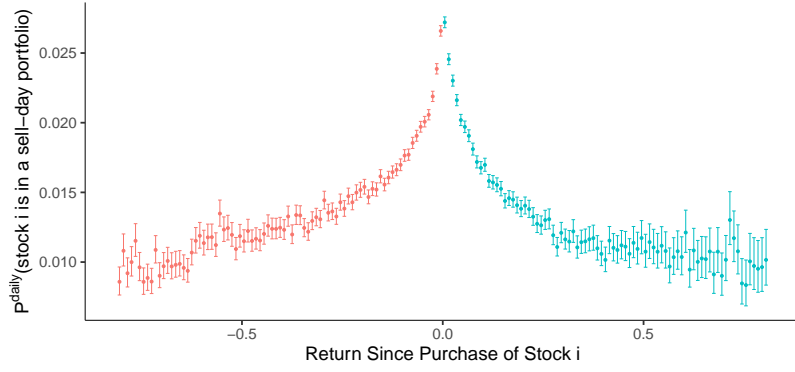


Figure A8: Proportion of stocks being included in a sell-day portfolio (Daily portfolio sample). The bin-width is 0.01. The error bars are 95% confidence intervals.

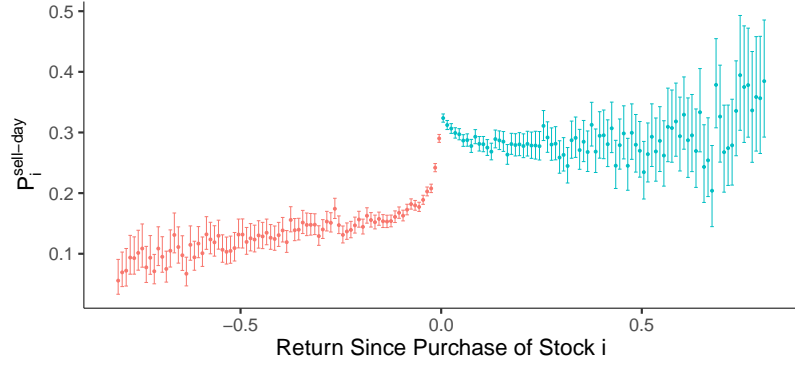


Figure A9: Proportion of stocks being sold ($P_i^{sell-day}$; Sell-day portfolio sample). The bin-width is 0.01. The error bars are 95% confidence intervals.

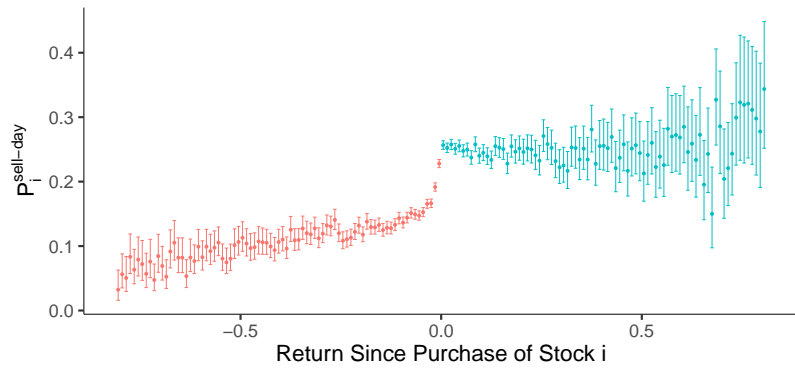


Figure A10: Proportion of stocks being sold ($P_i^{sell-day}$; Sell-day portfolio sample given $N \geq 2$). The bin-width is 0.01. The error bars are 95% confidence intervals.



Figure A11: Number of stocks in a portfolio on return since purchase (Daily portfolio sample)



Figure A12: Number of stocks in a portfolio on return since purchase (Daily portfolio sample given $N \geq 2$)

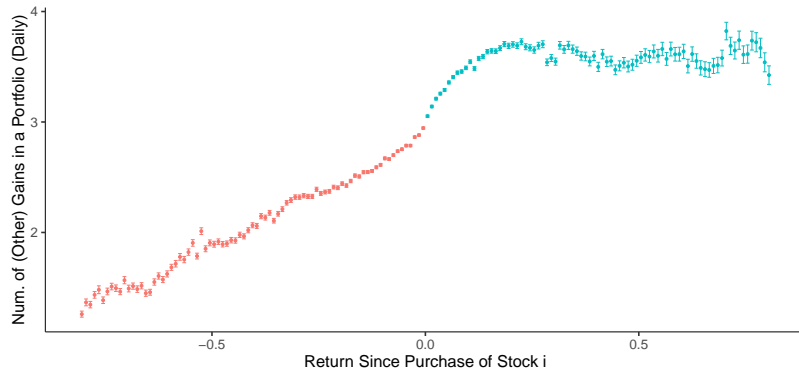


Figure A13: Number of gains in a portfolio on return since purchase (Daily portfolio sample given $N \geq 2$). One gain was not counted in the gain domain.

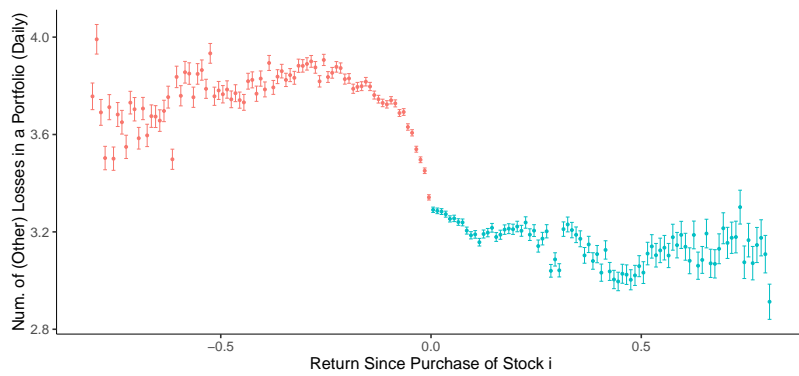


Figure A14: Number of losses in a portfolio on return since purchase (Daily portfolio sample given $N \geq 2$). One loss was not counted in the loss domain.



Figure A15: Number of stocks in a portfolio on return since purchase (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$)

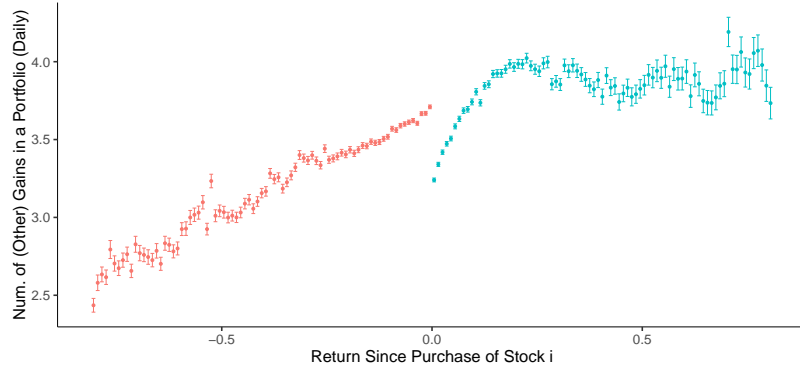


Figure A16: Number of gains in a portfolio on return since purchase (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). One gain was not counted in the gain domain.



Figure A17: Number of losses in a portfolio on return since purchase (Daily portfolio sample given $NG \geq 1$ and $NL \geq 1$). One loss was not counted in the loss domain.

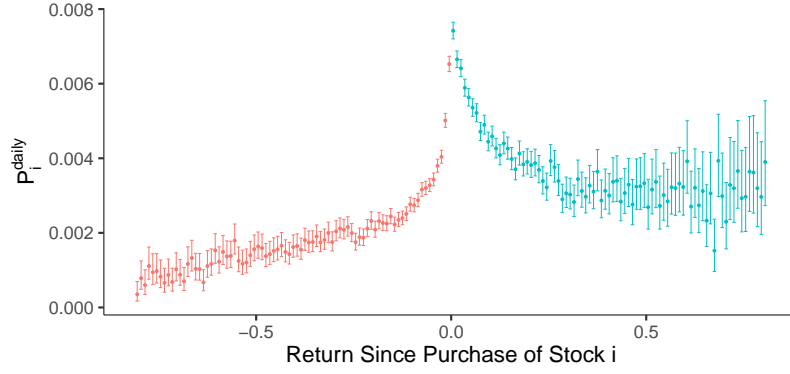


Figure A18: Proportion of stocks being sold (P_i^{daily} ; Daily portfolio sample given $N \geq 2$). The bin-width is 0.01. The error bars are 95% confidence intervals.

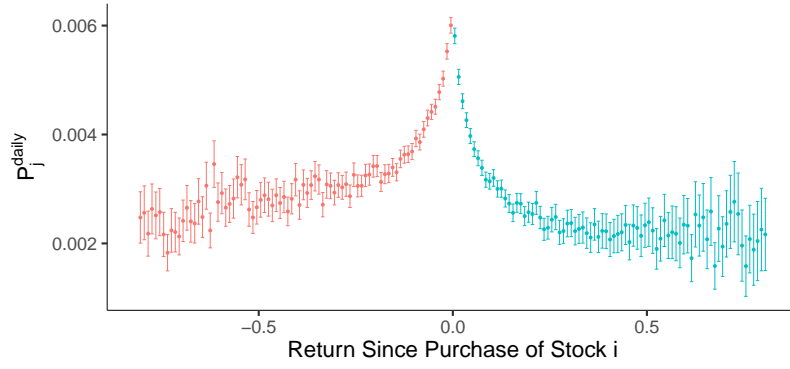


Figure A19: Proportion of other stocks in the portfolios being sold (P_j^{daily} ; Daily portfolio sample given $N \geq 2$). The bin-width is 0.01. The error bars are 95% confidence intervals.

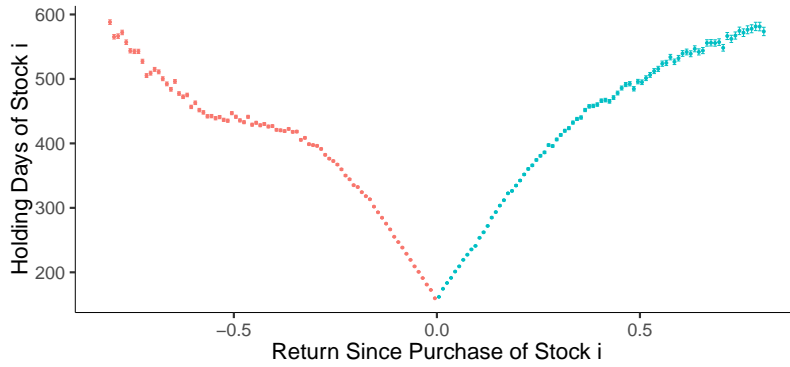


Figure A20: Holding Days of Stock i (Daily portfolio sample given $N \geq 2$). The bin-width is 0.01. The error bars are 95% confidence intervals.

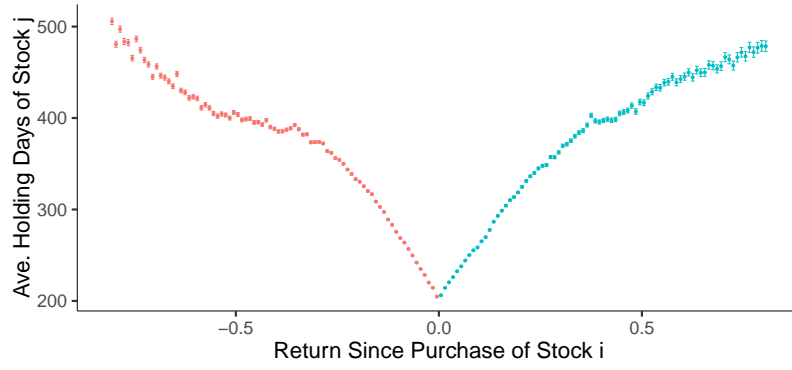


Figure A21: Average Holding Days of Stock j (Daily portfolio sample given $N \geq 2$). The bin-width is 0.01. The error bars are 95% confidence intervals.

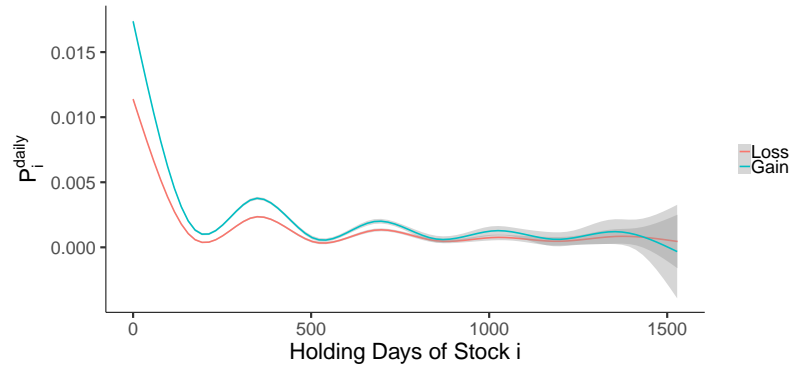


Figure A22: P_i^{daily} on Holding Period (Daily portfolio sample).

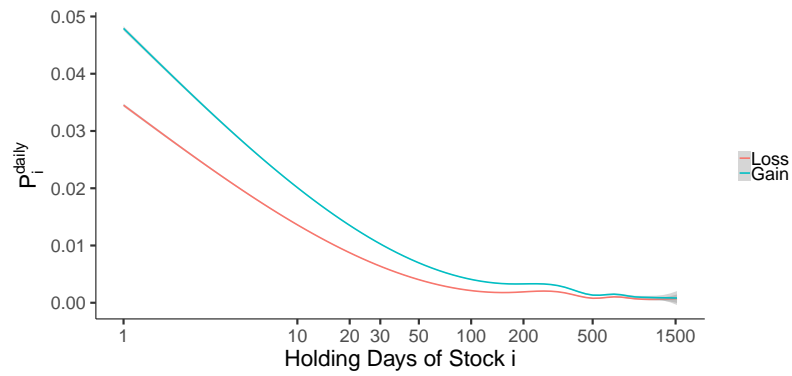


Figure A23: P_i^{daily} on Holding Period (Daily portfolio sample). The x-axis are log transformed.

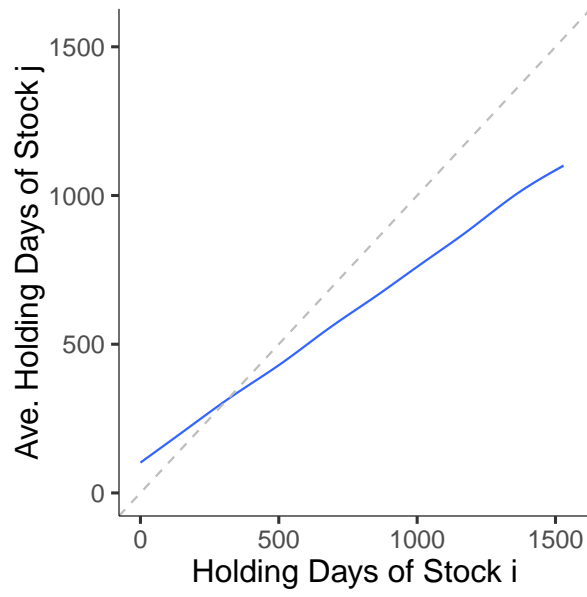


Figure A24: Holding Periods of Stock i and Average Holding Period of Stocks j .

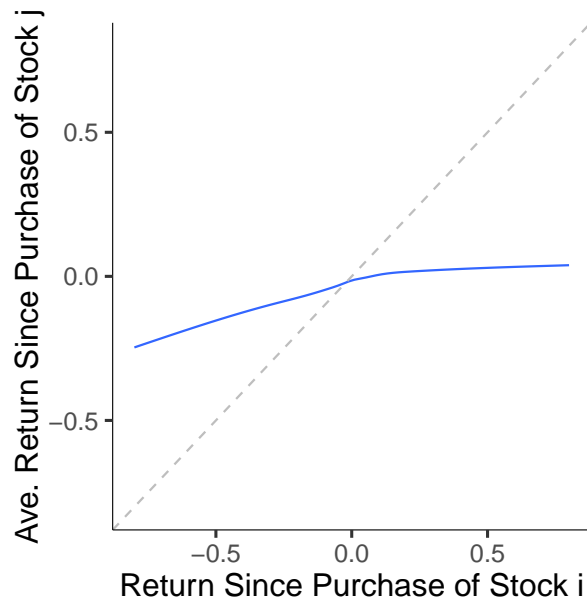


Figure A25: Return of stock i and average return of stocks j (Daily portfolio sample given $N \geq 2$).

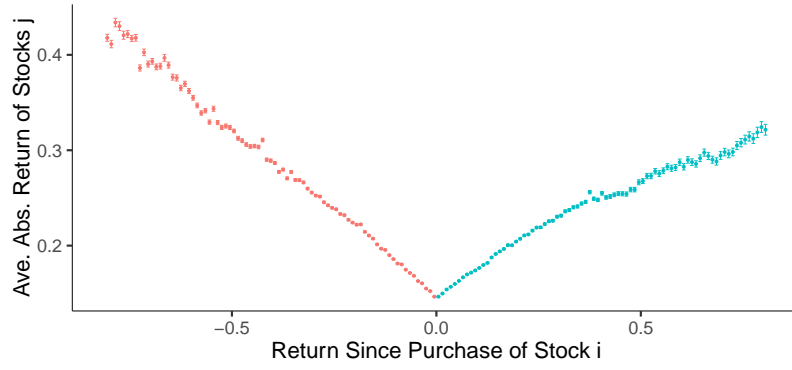


Figure A26: Return of stock i and average absolute return of stocks j (Daily portfolio sample given $N \geq 2$).

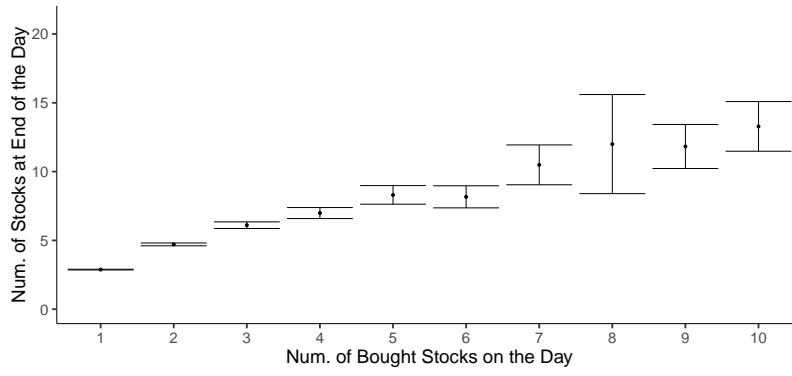


Figure A27: Number of Stocks Bought on the Day and Number of End-Day Stocks (Buy-day portfolio sample given $N \leq 10$).

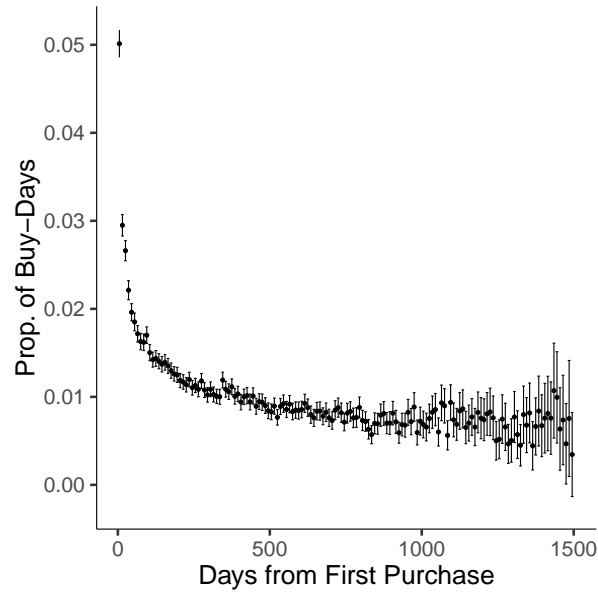


Figure A28: Proportion of Buy-Days among All days as a Function of Distance from a First Buy Day (Daily portfolio sample).

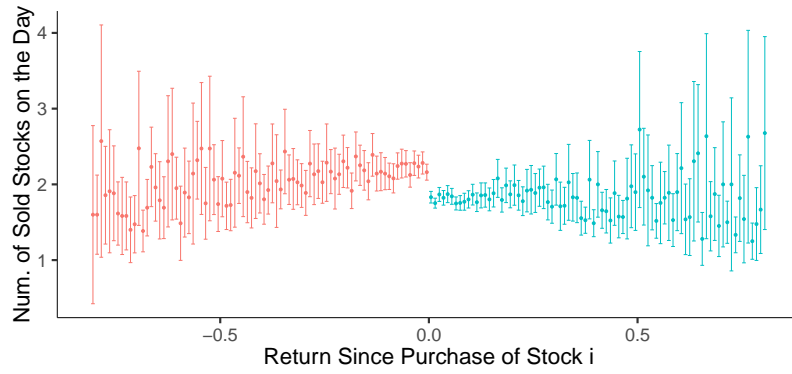


Figure A29: Number of sold stocks on the day given that stock i is sold.

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.
- 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.