Selling Winners or Losers: Two-Stage Decision Making and the Disposition Effect in Stock Trading

Hiroaki Sakaguchi, Neil Stewart, Lukasz Walasek

Abstract

Current methods for estimating the disposition effect implicitly assume that all stocks are evaluated simultaneously in a single decision stage. Here we propose a two-stage model where investors first decide whether to sell a stock in the domain of gains or losses, and only then choose a stock to sell from within their chosen domain. As evidence, we show that the probability of individual gains being sold is inversely proportional to the number of gains in the portfolio, but is not associated with the number of losses. Similarly, the probability of individual losses being sold is inversely proportional to the number of losses in the portfolio, but is not associated with the number of gains. There are two consequences for the disposition effect: First, sell decisions are about the domain of gains versus losses, not just about individual stocks. Second, current regression methods must be refined to avoid substantial bias.

Keywords: disposition effect, gains, losses, decision making, loss aversion

14 October 2017. Email neil.stewart@wbs.ac.uk

This work was supported by Economic and Social Research Council grants ES/K002201/1, ES/N018192/1, and Leverhulme grant RP2012-V-022.
1 Introduction

One of the most well-evidenced behavioral biases in finance is the disposition effect, in which people are more likely to sell stocks that have gained value since they bought them than stocks that have lost value (Odean, 1998; Shefrin & Statman, 1985). More generally, the idea that people treat gains and losses differently is well established in psychology and economics (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991) and is embodied in the concept of loss aversion which is core in behavioral economics (Camerer, 2005). The idea that outcomes are evaluated against a reference level, often taken to be the status quo, is also well established (Kőszegi & Rabin, 2006, 2007; Lopes & Oden, 1999). This paper shows that, in the domain of finance, whether a stock has gained or lost value is psychologically elemental, such that (a) stocks are grouped together, with decisions taken about selling at this category level without reference to the magnitudes of the gains and losses, and (b) stocks in gain are considered separately from stocks in loss, and vice versa. As such we demonstrate that the disposition effect is, at least in the large part, a gain-loss-domain-level effect and not only an individual-stock-level effect.

The disposition effect is robust, and has been demonstrated using real stock trading data for private investors (Brown et al., 2006; Grinblatt & Keloharju, 2000; Odean, 1998), professional traders (Garvey & Murphy, 2004) and in laboratory experiments (Weber & Camerer, 1998). In many studies, the magnitude of the disposition effect is estimated using data from days upon which at least one stock is sold (sell-day portfolios) (e.g., Kaustia, 2010; Odean, 1998). Regression models are used to estimate the probability that a particular stock is sold while controlling for other properties of a given stock (e.g., return since purchase, price volatility, holding period). The disposition effect is observed when the probability that a stock is sold is higher when it is in gain than when it is in loss, other things being equal. The disposition effect is substantial. For example, Odean (1998) reported that gains are, on average, 1.5 times as much likely to be sold as losses in the US retail investors’
portfolios. Similarly, Kaustia (2010) showed that gains are twice as much likely to be sold as losses when Finish investors sold a stock with a short holding period.

Here we propose that existing methods for estimating the disposition effect are inadequate because they make an erroneous assumption about the decision processes of individual investors. Consider an investor who is trying to decide on which stock to sell from his or her portfolio. The investor may consider all of the stocks in the portfolio, comparing their past performance and trying to predict their future outlook. If this investor exhibits the disposition effect, his or her decision will be swayed towards selling a stock in gain over a stock in loss. In this account, whether a stock is in gain or in loss is just one of many features used to assess each individual stock. This decision rule aligns well with the assumptions of the regression techniques used to estimate the disposition effect, in which the probability of each stock being sold is estimated simultaneously across domains of gains and losses. We refer to this approach as the one-stage model to reflect its implicit assumptions about the investors’ decision process. Now, consider an alternative process in which an investor seeks to minimize the cognitive cost associated with the complex trade-offs of comparing stocks in gain with stocks in loss. Our investor therefore begins by answering a simple but important question: Do I sell a stock in gain, or do I sell a stock in loss? This decision is exogenously made without any consideration of individual stocks in the portfolio and its composition, but can be influenced by the investor’s tendency to sell gains over losses. In the second stage of the decision process, the investor is left with one of two possible choice contexts: If he or she initially decided to sell a gain then he or she must now decide which gain to sell. Alternatively, if he or she initially decided to sell a loss, he or she must now decide which loss to sell. We refer to this process as the two-stage model, since the investors begin by selecting a domain from which they will sell in the first stage, and only then in the second stage do they evaluate the subset of stocks in the domain they chose in the first stage.

In Section 2 we argue that, from a psychological perspective, the two-stage model offers a plausible account of the investors’ decision process. We propose that, in order to reduce decision complexity, people segregate outcomes into gains and losses and engage only in within-domain
comparisons when evaluating individual stocks. In Section 3 we describe the Barber and Odean (2000) data set which we use to estimate the models. In Section 4 we outline the unique predictions that this two-stage model makes about how the size of the disposition effect should vary with the number of gains and losses in a portfolio. In Section 5 we show how the disposition effect is sensitive to the number of gains and losses in a portfolio, exactly as the two-stage model predicts. In Section 6 we show the implications for using regression models to estimate the disposition effect. In Section 7 we consider what the evidence for the two-stage model means for the origins of the disposition effect.

2 The Psychology of a Decision to Sell

When faced with a complex choice problem, individual decision makers tend to adapt and choose strategies that reflect a trade-off between decision accuracy and the cognitive cost of deciding—satisficing (Simon 1955, 1956). As a result, decision makers are likely to use sequential and non-compensatory decision rules (Gigerenzer & Gaissmaier, 2011; Payne, Bettman, & Johnson, 1993). A common feature of these strategies is that people attempt to reduce size of a choice set (i.e., the number of alternatives in consideration) using a single criterion at a time (Brandstätter et al., 2006; Tversky, 1972). Such models stand in stark contrast with the standard economic view, in which all available information is considered in making a decision. Here we focus on people’s tendency to reduce the complexity of a decision context by first segregating choice objects into the domains of gains and losses, and then subsequently evaluating the options available within a domain.

The distinction between the positive and the negative is reflected in the psychological theories of language, attitude formation, attention allocation, reinforcement learning, and decision making (Cacioppo & Berntson, 1994; Rozin & Royzman, 2001). At the most rudimentary level, the common assumption is that people perceive different alternatives as advantages or disadvantages relative to some neutral reference point, often given by the current status quo (Kőszegi & Rabin, 2006; Novemsky & Kahneman, 2005; Thaler & Johnson, 1990). The majority of the existing work related to judgment and decision making under risk and uncertainty focused solely on the asymmetric
weighting of gains and losses. It is widely accepted that the anticipated negative emotions associated
with a loss are stronger than the anticipated positive emotions associated with a gain of an equal
magnitude (Kermer, Driver-Linn, Wilson, & Gilbert, 2006). Demonstrations of such loss aversion (or
negativity bias) apply to both monetary and nonmonetary domains (Baumeister, Bratslavsky,
Finkenauer, & Vohs, 2001). The existence of the disposition effect has been interpreted as support of
the asymmetric weighing of gains and losses. People may avoid realizing a loss out of a concern for
the intensity of negative feelings, and hence become more likely to sell a stock in gain (Weber &
Camerer, 1998). Here the label of loss aversion is a purely descriptive concept that does not come
with any assumptions about the underlying decision processes. The goal of the present work is to flesh
out the decision rule that may underpin the decision to sell a gain or a loss by individual investors.

Psychological theory suggests that when people evaluate anticipated feelings of positive and
negative events, they naturally engage in a within-domain comparison (Kahneman & Miller, 1986;
McGraw et al., 2010). That is, people may choose whether a given situation or an outcome falls into
category of gains and losses, and only then proceed to compare it with outcomes within the same
domain. Such mechanisms are also consistent with many models of relative judgment, where the
relative comparison context is often constructed by separating outcomes using a natural anchor, such
as zero point or other neutral value (Marsh & Parducci, 1978; Parducci, 1983; Stewart, Chater, &
Brown, 2006). For example, consider a situation in which you are trying to form an evaluative
judgment about losing your baggage at an airport. In forming a relevant set of comparable events, one
is likely to think of other negative things that might have happened in the past whilst traveling. It is
unlikely however, that one would bring to mind both negative and positive events in order to evaluate
an unpleasant experience. Studies of loss aversion in risky choice support the idea that gains are
evaluated only against other gains and that losses are evaluated only against other losses. For
example, Walasek and Stewart (2015) showed that people’s reluctance to accept mixed lottery
gambles is largely dependent on the ranges of possible gains and losses that people store in their
memory. In their experiments, they found that people exhibit no loss aversion for symmetric 50-50
gambles as long as the gain appears attractive relative to other gains, and when the loss appears small relative to other losses. This strong context-sensitivity of loss aversion could not occur if people were making across-domain comparisons.

In sum, whether something is a gain or a loss is a psychologically salient category. Here we propose that complex decisions such as choosing which stock to sell rely on a separation of stocks that are in gains from those that are in loss. Additionally, we suggest that stocks in gains will be compared with others in gain, while stocks in loss will be compared with others in loss.

3 Data

Our data are historical stock transactions for individual investors in the US. The trades were completed through a large discount brokerage between January 1991 and November 1996. These data were previously used in studies of disposition effect by Barber and Odean (2000, 2001, 2002) and Hartzmark (2015). We merged trades with the historical prices retrieved from the Center for Research in Security Price (CRSP). Because the purchase prices of stocks bought before the beginning of the transaction data are unknown, we excluded all accounts which had positions at the end of January 1991 so that we have complete price data for all portfolios. Multiple intra-day trades conducted by the same investor on the same stock were aggregated with quantity weighted prices. We extracted sell trades which changed a net position from positive to non-negative (i.e., sell trades leading to short positions were excluded), and reconstructed the portfolios held by the corresponding accounts on these sell dates (sell-day portfolios). Short positions and positions opened on sell days were excluded from the remaining portfolios. The return since purchase was calculated by using a quantity weighted average purchase price of a stock for a given account and a closing price of the stock as of one day prior to the sell date. Commissions and dividends were not included in the calculation of returns. If a sell-day portfolio contained one or more stocks with missing variables in either the CRSP data or the transaction data, the whole sell-day portfolio was excluded. Because of this portfolio-based rather than stock-based exclusion, the compositions of all sell-day portfolios in our sample were exactly the
same to those in actual investors’ portfolios. (Note that we also conducted the analysis on the sample based on the stock-base exclusion. The results are nearly identical to those reported in this paper.) Because we are interested in the investors’ choice of stock for sale, we extracted sell-day portfolios consisting of two or more stocks. Sell-day portfolios consisting of only gains or only losses and those including stocks at a zero return were excluded. Further we extracted sell-day portfolios where exactly one stock was sold. These one-sale portfolios are 84.5% of all sell-day portfolios. The summary statistics for the portfolios used are presented in Tables A1, A2, and A3 in Appendix 1.

Table 1 summarizes the notation we use in the analysis. \( N_G \) is the number of gains in a portfolio and \( N_L \) is the number of losses, so that the total number of stocks in a portfolio is \( N_{G+L} = N_G + N_L \). \( Sell, Gain, \) and \( Loss \) are 0/1 dummy variables indicating whether a stock is sold, and whether it is in gain or loss. \( P(Gain) \) is the average value of \( Sell \) over the gains in portfolios. As such it represents the probability that a single individual stock in gain is sold. Analogously, \( P(Loss) \) represents the probability that a single individual stock in loss is sold. Because we select only the sell-day portfolios where exactly one stock was sold, \( P(Gain) \times N_G + P(Loss) \times N_L = 1 \).

The measure of the disposition effect at the level of individual stocks is \( \beta = P(Gain)/P(Loss) \). The measure of the disposition effect at the gain-loss domain level is \( \beta = \frac{[P(Gain) \times N_G]}{[P(Loss) \times N_L]} = \beta \frac{N_G}{N_L} \).

We also include control variables in our multivariate analyses for the return since purchase, \( Return \), the return in the past 20 days, \( Return_{20} \), the volatility in the past 20 days, \( Volatility_{20} \), and the number of days that a stock has been held for, \( Holding \ Days \).
Table 1. Notations and Descriptions of Variables Used in Sections 4, 5, and 6

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_G$</td>
<td>The number of gains in a sell-day portfolio</td>
</tr>
<tr>
<td>$N_L$</td>
<td>The number of losses in a sell-day portfolio</td>
</tr>
<tr>
<td>$N_{G+L}$</td>
<td>The total number of stocks in a sell-day portfolio</td>
</tr>
<tr>
<td>$Sold$</td>
<td>A dichotomous variable having a value of 1 if the stock was sold otherwise 0</td>
</tr>
<tr>
<td>$Gain$</td>
<td>A dichotomous variable having a value of 1 if the stock was in gain, otherwise 0</td>
</tr>
<tr>
<td>$Loss$</td>
<td>A dichotomous variable having a value of 1 if the stock was in loss, otherwise 0</td>
</tr>
<tr>
<td>$P(Sold)$</td>
<td>The probability that an individual stock is sold. For empirical data, this is a proportion</td>
</tr>
<tr>
<td>$P(Gain)$</td>
<td>The probability that an individual gain is sold. For empirical data, this is the average of $Sold$ over gains</td>
</tr>
<tr>
<td>$P(Loss)$</td>
<td>The probability that an individual loss is sold. For empirical data, this is the average of $Sold$ over losses</td>
</tr>
<tr>
<td>$\beta$</td>
<td>The size of the disposition effect at the individual-stock level (hence lower-case $\beta$)</td>
</tr>
<tr>
<td>$B$</td>
<td>The size of the disposition effect at the gain-loss-domain level (hence capital $B$)</td>
</tr>
<tr>
<td>$Return$</td>
<td>The stock’s return since purchase</td>
</tr>
<tr>
<td>$Return_{20}$</td>
<td>The stock’s return for 20 days prior the sell day</td>
</tr>
<tr>
<td>$Volatility_{20}$</td>
<td>The stock’s volatility for 20 days prior the sell day</td>
</tr>
<tr>
<td>$Holding\ Days$</td>
<td>The number of days that the stock has been held for, from first purchase to the sell day</td>
</tr>
</tbody>
</table>
4 Model Predictions

Taken together, the psychological literature suggests that investors may seek to simplify their decision by first choosing whether to sell a stock in gain or in loss and then choosing a particular stock from within either the domain of gains or losses. We contrast this two-stage choice with a single-stage choice where people compare all stocks (gains and losses) simultaneously to choose which stock to sell. The two-stage model offers unique predictions about the relationship between the portfolios’ composition and the size of the disposition effect. If traders follow the two-stage model and choose the domain of either gains or losses before deciding which stock to sell, we should find that the probability of an individual gain being sold is sensitive to the number of gains in the portfolio, but not the number of losses. Similarly, the probability of an individual loss being sold should be sensitive to the number of losses in the portfolio, but not the number of gains. This prediction follows from the two-stage model because once a trader decides upon a domain (either gains or losses) only stocks within that domain will be in competition to be sold. In the one-stage model, on the other hand, the probability that a gain or a loss is sold will be with a function of all stocks in the portfolio.

To preempt the results in Section 5, our analysis shows that a mixture of the conventional one-stage model and the proposed two-stage model are required to fit the data. The implications are profound. First, this indicates that people take a category-level decision about whether to sell a gain or a loss independently of the properties of the individual stocks involved. This means that the disposition effect is, in large part, a portfolio-level phenomena rather than merely an individual-stock level phenomena. The second implication is that the current regression approaches to estimating the disposition effect must be corrected, which we address in Section 6.

4.1 The one-stage model

The conventional method for estimating the selling probability of individual stocks assumes that investors evaluate all stocks in their portfolio simultaneously to choose one stock to sell. The disposition effect is at the individual stock level. We consider the measure of the individual-stock-
level disposition effect $\beta$ which is the single free parameter in the one-stage model, and reflects the relative probabilities of selling an individual single gain, $P(Gain)$ rather than an individual single loss, $P(Loss)$.

Because we take only sell-day portfolios where exactly one stock is sold, by definition, we use our constraint that $P(Gain) \times N_G + P(Loss) \times N_L = 1$ (see the data-selection criteria described in Section 3). We also have, from the definition of $\beta$, that $\beta = P(Gain)/P(Loss)$. Substituting for $P(Loss)$ gives

$$P(Gain) = \frac{1}{N_G + \frac{N_L}{\beta}}$$  \hspace{1cm} (1)

and substituting for $P(Gain)$ gives

$$P(Loss) = \frac{1}{\beta N_G + N_L}$$  \hspace{1cm} (2)

Thus, according to the one-stage model, both $P(Gain)$ and $P(Loss)$ should be sensitive to both $N_G$ and $N_L$. This is because investors are assumed to evaluate all of the stocks in a portfolio simultaneously, across both gains and losses.

Figure 1 shows the one-stage model predictions for $P(Gain)$ and $P(Loss)$ as a function of $N_G$ and $N_L$ using the best-fitting value of $\beta = 2.16$ (see Appendix 4). Otherwise, the selection of a stock for sale is assumed to be random. In Figure 1A, the x-axis represents $N_G$ and each line represents different value of $N_L$. Given a fixed $N_L$ (i.e., for a given line), the larger $N_G$ the smaller $P(Gain)$. When $N_L = 1$, the curve in $N_G$ will be relatively steep because changes of $N_G$ in the denominator are large compared to $N_L/\beta$. Effectively, each extra gain in the portfolio takes a large share of the total probability of selling because the losses are taking only one share. When $N_L = 5$, the curve in $N_G$ will be relatively flat because changes of $N_G$ in the denominator are small compared to $N_L/\beta$. Effectively, each extra gain in the portfolio takes only a small share of the total probability of selling because the losses are taking five shares. Figure 1B replots Figure 1A exchanging the roles
of $N_G$ and $N_L$, which is useful for comparison with plots of the data later. Figure 1C shows $P(Loss)$ as a function of $N_G$ and $N_L$. This is just a transformation of Figure 1A, but is useful later too. The curves are similar to those for $P(Gain)$ seen in Figure 1A. Figure 1D replots Figure 1C exchanging the roles of $N_G$ and $N_L$.

Figure 1. $P(Gain)$ and $P(Loss)$ as a function of $N_G$ and $N_L$ in the one-stage model. The right panels replot the data, swapping the roles of $N_G$ and $N_L$.

4.2 The two-stage model

In the two-stage model, investors first choose whether to sell from the gain domain or the loss domain, before then choosing a specific stock from their chosen domain. The disposition effect is at the level of the gain-loss domain, and not at the level of individual stocks as it is in the one-stage model.
model. In the two-stage model the single free parameter is the domain-level disposition effect $B$ that is our free parameter.

Thus, we can begin again with our constraint that $P(Gain) \times N_G + P(Loss) \times N_L = 1$, but instead of substituting for $\beta$ we substitute for $B = \frac{[P(Gain) \times N_G]}{[P(Loss) \times N_L]}$ to get

$$P(Gain) = \frac{1}{N_G \left(1 + \frac{1}{B}\right)}$$

and

$$P(Loss) = \frac{1}{N_L \left(1 + B\right)}$$

It is obvious from these equations that $P(Gain)$ depends only on $N_G$ and that $P(Loss)$ depends only upon $N_L$ in the two-stage model. Figure 2 shows the two-stage model predictions for $P(Gain)$ and $P(Loss)$ as a function of $N_G$ and $N_L$. We use best-fitting value of $B = 2.04$ (see Appendix 4). Figures 2A and B show that $P(Gain)$ is inversely proportional to $N_G$ but independent of $N_L$. Figures 2C and D show that $P(Loss)$ is inversely proportional to $N_L$ but independent of $N_G$. 
Figure 2. $P(Gain)$ and $P(Loss)$ as a function of $N_G$ and $N_L$ in the two-stage model. The right panels replot the data, swapping the roles of $N_G$ and $N_L$.

To recap, if the one-stage model is correct, then $\beta$ should be a constant. This means that $P(Gain)$ and $P(Loss)$ will be a function of the total number of gains and losses in the portfolio, $N_G + N_L$. If the two-stage model is correct, then $B$ should be a constant. This means that $P(Gain)$ should be inversely proportional to $N_G$, and $P(Loss)$ should be inversely proportional to $N_L$.

5 Results

Below we test how the disposition effect is sensitive to the composition of the portfolio. First, we present some simple descriptive statistics of portfolios in our data, the disposition effect.
Then, we explore how the disposition effect varies with portfolio composition. To preempt the findings, we see a pattern that looks remarkably like the signature from the two-stage model described above.

5.1 The disposition effect at individual stock level

First, we confirmed the presence of disposition effect in the data. Figure 3 compares $P(Gain)$ (the grey bar) with $P(Loss)$ (the red bar), showing that individual gains are on average about 1.8 times as much likely to be sold as individual losses (i.e., $\beta = 1.8$).

![Figure 3. The disposition effect. The error bars are 95% bootstrapped confidence intervals computed with 1,000 resamples, corrected for clustering by accounts and sell dates.](image)

5.2 Composition-sensitivity of the disposition effect

The degree of the disposition effect depends on the composition of sell-day portfolios. In Figure 4, sell-day portfolios were divided into four bins depending on the ratio of the number of gains and losses in the portfolio. The size of the disposition effect (i.e., the difference between adjacent grey and red bars) reduces considerably as the ratio of the number of gains to the number of losses increases. For the Mostly Losses bin, individual gains are on average about 3.8 times more likely to be sold as individual losses, much larger than the 1.8 times for all portfolios. For the Mostly Gains
bin, the disposition effect reverses such that losses are now more likely to be sold than gains. This very simple calculation of proportions is complemented with a multivariate analysis of composition sensitivity in Appendix 2, where we control for the returns, number of days held, and volatility of individual stocks, and include fixed effects for account and stock-by-date. The multivariate analysis confirms the pattern seen in Figure 4.

![Bar chart showing the disposition effect in different portfolio compositions](image)

Figure 4. The disposition effect depends on the composition of the portfolio. The error bars are 95% confidence intervals computed with the bootstrap method with 1,000 resamples, corrected for clustering by accounts and sell dates.

5.3 Within-domain sensitivity

Here we show that the probability of an individual gain being sold is sensitive mostly to the number of gains but not the number of losses. And the probability of an individual loss being sold is sensitive mostly to the number of losses but not the number of gains. We call this the within-domain sensitivity. Figure 5A plots the proportion of sales taken by an individual gain, $P(Gain)$, as a function of the numbers of gains and losses in the portfolio, $N_G$ and $N_L$. Figure 5B replots these data,
swapping the roles of $N_G$ and $N_L$. These two panels make it visually obvious that for the probability that an individual gain is sold, $N_G$ has a large effect while $N_L$ has, at most, only a small effect. $P(Gain)$ is nearly inversely proportional to $N_G$, but is unrelated to $N_L$. Figures 5C and D repeat these plots for the proportion of losses sold, $P(Loss)$. Now the pattern is reversed, with $N_G$ having, at most, only a small effect while $N_L$ has a large effect. $P(Loss)$ is nearly inversely proportional to $N_L$, but is unrelated to $N_G$.

![Graphs showing probability distributions](image)

Figure 5. $P(Gain)$ and $P(Loss)$ as a function of $N_G$ and $N_L$ in the empirical data. The shaded areas are bootstrapped 95% confidence intervals, with clustering by accounts and sell dates. The right panels replot the data, swapping the roles of $N_G$ and $N_L$.

The plot of the empirical data in Figure 5 bears a striking resemblance to the two-stage model predictions in Figure 2. To repeat the argument in Section 4.2, this within-domain sensitivity
follows quite trivially from the two-stage model. Once the domain of gains is chosen for selling, the probability that any individual gain is sold is proportional to $1/N_G$, assuming the specific gain to be sold is selected at random. And once the domain of losses is chosen for selling, the probability that any individual loss is sold is simply proportional to $1/N_L$, assuming the specific loss to be sold is selected at random. Note that, for a robustness check, the Appendix 3 replicates Figures 4 and 5, using the sample of tax-exempt accounts (see Figures A1 and A2).

5.4 Estimating a mixture of the one-stage and two-stage models

In order to estimate what proportion of sell-day portfolios in our data follow the two-stage model, we conducted an optimization. The optimization finds what mixture probability of the one-stage and the two-stage models best fits the sell-day portfolios (see Appendix 4 for details). The results show that the composition of the optimized model is 43%, 95% CI [35%, 56%] of the one-stage model and 57%, 95% CI [44%, 65%] of the two-stage model. In the one-stage model, the individual-stock level disposition effect $\beta = 2.08$, 95% CI [1.12, 4.12], which represents individual gains being about 2.08 times more likely to be sold than individual losses. In the two-stage model, the domain-level disposition effect $\beta = 2.09$, 95% CI [1.12, 3.21], which represents the gain domain is 2.09 times more likely than the loss domain to be chosen in the first stage. The optimized model fits the empirical data better than the one-stage model alone and the two-stage model alone.

6 Implications for Regression-Based Estimates of the Disposition Effect

We have shown that the disposition effect is strongly related to the composition of the portfolio. This within-domain sensitivity strongly implicates a two-stage model where an initial decision is taken about which domain to make a sale from before individual stocks are considered. It therefore follows that regression models estimating the probability of individual stocks being sold should properly control for the composition of a portfolio. This does not mean that the two stages of the decision process need to be implemented in the regression framework. Instead, with an appropriate control for the composition of a portfolio we can account for the within-domain
sensitivity of the two-stage model. As a simple illustration, we compare four logistic models which differ from one another only in the way in which they control for $N_G$ and $N_L$. The dependent variable which is common for all four models is the log-odds of the decision to sell a stock ($Sell$). The common covariates are: $Gain$, $Gain \times Return$, $Loss \times Return$, $\sqrt{Holdings\ Days}$, $Gain \times Return \times \sqrt{Holdings\ Days}$, $Gain \times Return_{20}$, $Gain \times Volatility_{20}$, $Loss \times Return_{20}$, and $Loss \times Volatility_{20}$. See Table 1 for the description of these variables. The models differ only in controls for $N_G$ and $N_L$. Model 1 does not control for $N_G$ and $N_L$. Model 2 controls for the reciprocals of the total number of stocks across gains and losses in a portfolio interacting separately with the gain and the loss domain: \[ \left( \frac{1}{N_{G+L}} \times Gain \right) \] and \[ \left( \frac{1}{N_{G+L}} \times Loss \right) \]. Model 2 is very similar to the one-stage model, differing only in controlling for returns, volatility, and holding duration, and in having the logit link function as a logistic regression rather than directly modeling the probability of a sell. Model 3 includes separate interactions for the number of gains and for the number of losses: $(N_G \times Gain)$, $(N_L \times Loss)$, $(N_G \times Loss)$, and $(N_L \times Gain)$. Model 4 includes separate interactions for the reciprocals of the numbers of gains and losses: \[ \left( \frac{1}{N_G} \times Gain \right) \], \[ \left( \frac{1}{N_L} \times Loss \right) \], \[ \left( \frac{1}{N_G} \times Loss \right) \], and \[ \left( \frac{1}{N_L} \times Gain \right) \]. Model 4 is very similar to the mixture-model, differing only in controlling for returns, volatility, and holding duration, and in having the logit link function, and having the cross-domain interactions \[ \left( \frac{1}{N_G} \times Loss \right) \] and \[ \left( \frac{1}{N_L} \times Gain \right) \].

Figure 6 compares the model predictions for $P(Gain)$. The predictions of Model 4 (bottom row) are close to those seen in the empirical data in Figure 5. On the other hand, the predictions of Models 1, 2, and 3 in the first three rows of Figure 6 all deviate from the empirical data. Model 1 does not concern the composition of a portfolio at all, leading to the worst fit among four models. Model 2 controls for $(\frac{1}{N_{G+L}} \times Gain)$ and $(\frac{1}{N_{G+L}} \times Loss)$. However, this control is on the total number of stocks across gains and losses and thus cannot capture the within-domain sensitivity. The prediction is similar to that of the one-stage model seen in Figure 1. Model 3 captures the within-domain sensitivity
to some extent. However, as seen above, the relationship between \( P(Gain) \) and \( N_G \), and that between \( P(Loss) \) and \( N_L \) is not linear but inversely proportional. Therefore, the control in Model 3 is not sufficient. In summary, Figure 6 shows that a clear advantage of Model 4 over Models 1, 2 and 3.

That is, in order to capture the inverse proportionality within a domain, models should include the inverse of \( N_G \) and the inverse of \( N_L \) as control variables. Note that, because we used logistic models, Model 4 does not exactly control for the inverse proportional relationship between \( P(Gain) \) and \( N_G \).

However, Model 4 sufficiently captures a non-linear relationship to offer a good fit. Model 4 offers the highest \( R^2 \), and is preferred by AIC and BIC.

Table A6 in Appendix 5 presents the full regression estimates. We note two things here about the coefficients. First, the coefficient for the Gain dummy indicates the disposition effect, and varies considerably across the model specifications. Because Models 1-3 cannot capture the variation in the probability of selling a stock as a function of the number of gains and losses in the portfolio (compare the empirical effects in Figure 5 with the model predictions in Figure 6), there is substantial bias in the estimation of the disposition effect. Second, in Model 4, in line with the two-stage model and the within-domain sensitivity, the coefficient estimates for \( \left( \frac{1}{N_G} \times Gain \right) \) and \( \left( \frac{1}{N_L} \times Loss \right) \) (i.e., the within-domain effect) is much larger than those for \( \left( \frac{1}{N_G} \times Loss \right) \) and \( \left( \frac{1}{N_L} \times Gain \right) \) (i.e., the across-domain effect).
Figure 6. Comparison of the logistic regression model predictions for $P(Gain)$. Within each row, the right panels replot the data from the left panels, swapping the roles of $N_G$ and $N_L$. 
7 General Discussion

How do investors choose which stock to sell from their portfolio? We propose a two-stage decision rule, in which investors first decide whether to sell a stock in gain or a stock in loss without reference to the magnitudes of the gains or losses, and only then compare stocks individually within a given domain. This model is consistent with the existence of the disposition effect, but it also offers a new prediction about how the magnitude of this phenomenon should vary as a function of portfolio composition. More specifically, the two-stage model predicts that the probability of selling a particular stock in gain will depend only upon the number of stocks in gain in the portfolio and not the number of stocks in loss, and that the probability of selling a particular stock in loss will depend only upon the number of stocks in the portfolio in loss and not the number of stocks in gain. We tested this prediction using a large volume of stock trading data and found strong evidence for this within-domain sensitivity. Using a mixture model, we estimate that selling decisions are about a 50/50 mix of our two-stage model and the traditional one-stage model. Within-domain sensitivity must therefore be accounted for in the regression models used to estimate the disposition effect. We showed that a model in which we control for the reciprocals of the number of gains and losses in a portfolio offered a much better fit to the data.

7.1 Alternative explanations

We have deferred until now the discussion of other possible accounts of the composition-sensitivity of the disposition effect. For example, consider an alternative version of the two-stage model in which investors first evaluate each gain by comparing with other gains to identify one candidate gain to be sold. They also evaluate each loss by comparing with other losses and pick one candidate loss to be sold. Then, at the second stage, the candidate gain and the candidate loss are compared with one another and exactly one of them ends up being sold. Because the first stage of this alternative two-stage model requires only within-domain comparison, the model predicts the within-domain sensitivity and thus the composition-sensitivity of the disposition effect. But the interpretation
of the disposition effect in the model is the same as the two-stage model described earlier: the disposition effect is a gain-loss-domain-level bias, and not an individual-stock-level bias.

We can also consider a model where the evaluation for gains and that for losses are completely independent. Depending on exogenous factors (e.g., feeling), investors evaluate only gains on one day to decide whether to sell one of gains and evaluate only losses on another day to decide whether to sell one of losses. The disposition effect may be represented by a difference in the number of days where gains or losses are evaluated. Because, in this model, the evaluation is completely independent for each domain, the model predicts the within-domain sensitivity and the composition-sensitivity of the disposition effect.

Our data do not allow us to test the exact cognitive process through which investors select a stock to sell. Without process data, it is difficult to identify which model is more valid. Further research, perhaps using carefully controlled lab experiments, may be necessary to disentangle the exact origins of the portfolio-composition sensitivity.

Recently, Barberis and Xiong (2012) proposed a model of realization utility where investors derive utility by realizing either gains or losses from their portfolio. The within-domain sensitivity that we reported could be possibly explained with by the extended version of this model if we assume that investors who are holding more losses than gains are more likely to myopically realize (i.e., sell) one of the gains in order to compensate for the negative feelings associated by the presence of many losses. However, we have ruled out this alternative explanation because the propensity to realise a gain is not associated with the size of unrealized gains or losses in the portfolio as it is predicted by the realization utility model (see Appendix 6 for details).

7.2 The origin of the disposition effect

We consider how our two-stage model relates to the existing accounts of the disposition effect. While the origin of the disposition effect has been debated in the literature (Ben-David & Hirshleifer, 2012; Hens & Vlcek, 2011; Kaustia, 2010), there are three dominant explanations of the
effect: prospect theory and loss aversion; a belief in mean reversion; and regret-avoidance (Shefrin & Statman, 1985; Zuchel, 2010).

In the simplest form of explanation based on prospect theory, investors are assumed to have an s-shaped value function, while the reference point is determined by the original stock’s purchase price. The gains portion of the value function is concave while the losses portion of the value function is convex. Under these assumptions, investors evaluate an individual stock by integrating over their expectation of the stock’s future distribution of returns after transforming them with the s-shaped value function. Given a nearly symmetrical distribution of expected future returns, when the stock is in loss, a large part of the distribution of expected future returns is in the convex part of the value function, leading to investors being risk-seeking and thus to hold the stock. When the stock is in gain, a large part of the distribution is in the concave part of the value function, leading to investors being risk-averse and thus to sell the stock. In this way, the prospect theory explains the disposition effect at individual stock level. In this prospect theory explanation, the disposition effect emerges as individual stocks are evaluated according to prospect theory. It is harder to see how prospect theory might account for the domain-level disposition effect we observe. One might assume that people evaluate all in a domain stocks and integrate over them to get a domain level expectation, but this does rather defeat the non-compensatory motivation described in Section 2.

The mean-reversion account proposes that people hold a belief that a stock price is negatively autocorrelated and therefore a stock’s price should revert to a ‘long-term’ mean (Andreassen, 1987; Kahneman & Tversky, 1973). The belief in mean reversion suggests that stocks which have recently depreciated are likely to go up to reach the long-term mean, and conversely, stock which have recently appreciated are likely to go down towards the long-term mean. Consequently, investors tend to hold stocks in loss which are likely to be oversold relative to the long-term mean and to sell stocks in gain which are likely to be overbought relative to the long-term mean. In this explanation, investors’ decisions depend on how long the long-term is, how long they have held the stock, and how large the past price movement was. The latter two elements are individual
stock specific characteristics and cannot explain the domain-level disposition effect. Having said that, it may be possible that people believe that gains as a category turn to losses and that losses as a category turn to gains, regardless of the duration of holding days and the magnitude of the return of individual stocks. While such a belief would be closer to the first stage of our two-stage model, it is distinct from the decision rule assumed in the mean-reversion account.

The theory of regret avoidance (Bell, 1982; Loomes & Sugden, 1982) suggests that people anticipate feeling regret about their past decision of purchasing the stock when they consider realizing a loss on the stock, but anticipate feeling pride when they consider realizing a gain on the stock. Therefore, if people are on average regret-averse and pride-seeking, they are more likely to sell gains than losses. This is consistent with the previous finding that people are risk seeking in the loss domain until they must realize the loss, after which they are risk averse (Imas, 2016). The degree of a regret and the degree of a pride may depend on the magnitude of a loss or a gain. In this sense, the regret avoidance also assumes to operate on individual stock level. However, it is possible that people hesitate to realize a loss because they do not want to feel bad at all, regardless of how large is the realized loss is. Equally, they may prefer to realize a gain because they seek to feel proud regardless of the size of the realized gain. If so, the regret avoidance may not be influenced by the size of a stock’s return and may be based on a categorical thinking which conforms to the first stage of the two-stage model. People’s happiness may be insensitive to the size of a gain when they make an earning on an investment decision (Kassam et al., 2011), which supports the idea that people think in categorical terms and therefore their regret avoidance will result in the disposition effect at a portfolio level.

We do not offer a definitive psychological origin for the disposition effect. But the relationship we report between the disposition effect and the composition of a portfolio in terms of the number of gains and losses strongly implicates a two-stage approach where an initial gain-loss domain-level decision is also strongly contributing to the disposition effect. Thus, existing accounts must take into account the fact that people often take domain-level decisions about whether to sell a
winner or a loser. More pragmatically, our results show that the current methods for estimating disposition effect must be revised to account for the complexity of the portfolio composition sensitivity. Without such controls, the estimates of the magnitude of the disposition effect will be incorrect. Our results also indicate not just the primacy of gains and losses rather than absolute value in people’s decision making, but that the gain-loss category alone, without reference to magnitude, drives a substantial component of the sell decision.
### Appendix

#### A1 Summary Statistics for the Sell-Day Portfolios

**Table A1. Descriptive Statistics**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. sell-day portfolios</td>
<td>35,761</td>
</tr>
<tr>
<td>No. stocks</td>
<td>181,896</td>
</tr>
<tr>
<td>No. accounts</td>
<td>10,675</td>
</tr>
<tr>
<td>No. unique sell-dates</td>
<td>1,467</td>
</tr>
</tbody>
</table>

**Table A2. Summary of Control Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>25th pctile</th>
<th>Median</th>
<th>75th pctile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return since purchase</td>
<td>0.05</td>
<td>0.41</td>
<td>-0.11</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>Holding days</td>
<td>247</td>
<td>290</td>
<td>50</td>
<td>142</td>
<td>333</td>
</tr>
</tbody>
</table>

**Table A3. Percentage of Sell-Day Portfolios by Composition**

<table>
<thead>
<tr>
<th>Proportion of Sell-Day</th>
<th>Number of Losses in a Portfolio</th>
<th>Number of Gains in a Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Table A3" /></td>
<td></td>
</tr>
</tbody>
</table>
A2 A Multivariate Analysis of Composition Sensitivity in the Disposition Effect

In order to confirm the composition-sensitivity of the disposition effect in multivariate setting, a linear regression was conducted. The dependent variable is the dichotomous variable *Sell* taking the value of 1 if a stock was sold, otherwise 0. The independent variables are *Gain* × *Gain Loss Ratio Bin, Best, Worst, Gain × Return, Loss × Return, √Holding Days*, *Gain × Return × √Holding Days, Loss × Return × √Holding Days, Gain × Return*₂₀, and, *Gain × Volatility*₂₀. *Gain Loss Ratio Bin* includes four bins: Mostly Losses (*N_G:N_L = 1:2+1*), More Losses (*N_G:N_L = 1:2−1:1*), More Gains (*N_G:N_L = 1+:1−2:1*), and Mostly Gains (*N_G:N_L = 2+:1*). *Best* and *Worst* are dummies for the best and worst performing stocks in a sell-day portfolio. Hartzmark (2015) showed, the best and worst performing stocks in a portfolio are more likely to be sold than other middle performing stocks (the rank effect). Fixed effects of accounts and stock-by-dates were included. The standard errors were clustered by accounts and sell dates.

Table A4 reports the coefficients. The first four rows show the effect of *Gain* (i.e., the disposition effect) interacting with *Gain Loss Ratio Bin*. Comparing the coefficients and corresponding confidence intervals among the four bins, it is clear that the disposition effect decreases from Mostly-Losses-Bin (the first row) to Mostly-Gains-Bin (the fourth row), showing that the larger the number of gains relative to the number of losses in a portfolio the smaller the disposition effect. The results are consistent with the composition-sensitivity of the disposition effect seen in Figure 4.
Table A4. A Linear Regression for Composition-Sensitivity of the Disposition Effect

| IV | Coefficient | LL | UL | Clustered SE | t value | Pr(>|t|) |
|----|-------------|----|----|-------------|--------|---------|
| Gain × Mostly-Losses-Bin | 0.154 | 0.063 | 0.245 | 0.046 | 3.310 | 0.001 |
| Gain × More-Losses-Bin | 0.107 | 0.040 | 0.174 | 0.034 | 3.114 | 0.002 |
| Gain × More-Gains-Bin | 0.042 | -0.017 | 0.101 | 0.030 | 1.379 | 0.168 |
| Gain × Mostly-Gains-Bin | 0.017 | -0.046 | 0.079 | 0.032 | 0.522 | 0.602 |
| Best | 0.148 | 0.087 | 0.208 | 0.031 | 4.752 | 0.000 |
| Worst | 0.037 | -0.014 | 0.088 | 0.026 | 1.409 | 0.159 |
| √Holding days | -0.002 | -0.005 | 0.002 | 0.002 | -0.795 | 0.426 |
| Gain × Return | 0.199 | -0.152 | 0.550 | 0.179 | 1.111 | 0.267 |
| Gain × Return 20 | 0.066 | -0.206 | 0.338 | 0.139 | 0.473 | 0.636 |
| Gain × Volatility 20 | 0.001 | -0.006 | 0.007 | 0.003 | 0.167 | 0.867 |
| Gain × Return × √Holding days | -0.003 | -0.011 | 0.004 | 0.004 | -0.834 | 0.404 |
| Loss × Return × √Holding days | -0.011 | -0.031 | 0.009 | 0.010 | -1.040 | 0.298 |

R² = .935
Number of observations = 181,896

Note. Fixed effects of accounts and stock-by-dates were included. The Standard errors were corrected for clustering by accounts and sell dates.

A3 Robustness Check on Tax-Exempt Accounts

Investors might have tax motivations to realize a gain or realize a loss, and thus, might evaluate only gains or only losses in their portfolio on the sell day. For checking whether our findings are robust without tax-motivated investors, we repeated the analysis with a sample of tax-exempt accounts (i.e., IRA and Keogh accounts). The results are shown in Figures A1 and A2.

Figure A1 shows that the composition sensitivity of the disposition effect seen in Figure 4 is observed in the sample consisting of tax-exempt accounts.

Figure A2 shows $P(Gain)$ and $P(Loss)$ as a function of $N_G$ and $N_L$ on the sample of tax-exempt accounts. While portfolios with an extreme composition (e.g., portfolios consisting of one gain and five losses) tend to deviate the pattern seen in Figure 5, the within-domain sensitivity is mostly confirmed. That is, $P(Gain)$ is inversely proportional to $N_G$ but is not sensitive to $N_L$ and $P(Loss)$ is inversely proportional to $N_L$ but is not sensitive to $N_G$.

To recap, the findings of the main analysis are robust with the sample consisting of only tax-exempt accounts.
Figure A1. The disposition effect depends on the composition of the portfolio (tax-exempt accounts).

This figure corresponds to Figure 4 reducing the sample to observations for IRA and Keogh accounts.

The error bars are 95% confidence intervals computed with the bootstrap method with 1,000 resamples, corrected for clustering by accounts and sell dates.
Figure A2. \( P(Gain) \) and \( P(Loss) \) as a function of \( N_G \) and \( N_L \) in the empirical data (tax-exempt accounts). This figure corresponds to Figure 5 reducing the sample to observations for IRA and Keogh accounts. The shaded areas are bootstrapped 95% confidence intervals, with clustering by accounts and sell dates. The right panels replot the data, swapping the roles of \( N_G \) and \( N_L \).

**A4 Estimating the Mixture of One- and Two-Stage Models**

We estimate a mixture model, in which the probability that an individual stock is sold is a linear combination of the predictions of the one- and two-stage models. The one-stage model has free parameter \( \beta \) for the individual-stock-level disposition effect. The two-stage model has free parameter \( B \) for the domain-level disposition effect. We use free parameter \( \omega \) as the mixture parameter.

First, for each stock in sell-day portfolios, we calculated the probability of the stock being sold based on the one-stage model, \( P_{one}(Gain) \) and \( P_{one}(Loss) \), using Equations 1 and 2. We also
calculated the probability of stocks being sold based on the two-stage model, \( P_{two}(Gain) \) and \( P_{two}(Loss) \), using Equations 3 and 4. We combined the predictions for \( P(Gain) \) and \( P(Loss) \) across the one- and two-stage models using the mixture parameter \( w \).

\[
P_{mix}(Gain) = (1 - w)P_{one}(Gain) + wP_{two}(Gain)
\]

and

\[
P_{mix}(Loss) = (1 - w)P_{one}(Loss) + wP_{two}(Loss)
\]

We used the Nelder-Mead simplex algorithm to estimate values for \( \beta, \beta, \) and \( w \) by maximizing the likelihood of \( P_{mix}(Gain) \) and \( P_{mix}(Loss) \). To obtain 95% CIs for our parameter, estimates were bootstrapped using 1,000 samples, clustering our sampling by account and sell day.

Our best-fitting estimates are \( \hat{w} = 0.57, 95\% \ CI [0.44, 0.65], \hat{\beta} = 2.08, 95\% \ CI [1.12, 4.12], \) and \( \hat{\beta} = 2.09, 95\% \ CI [1.12, 3.21] \).

We also separately estimated the one-stage model and the two-stage model. For the one-stage model alone, the best-fitting \( \hat{\beta} = 2.16, 95\% \ CI [2.04, 2.30] \). For the two-stage model alone, the best-fitting \( \hat{\beta} = 2.04, 95\% \ CI [1.95, 2.14] \). These models fit less well than the mixture model. Table A5 reports the log-likelihood, AIC, and BIC for the one- and two-stage models and the mixture model. Figure A3 compares the predictions of the one- and two-stage models and the mixture model with the empirical data.

| Table A5. Model Selection Criteria for Three Optimized Models |
|-----------------|-----------------|-----------------|
| Model           | One-stage        | Two-stage        | Mixture         |
| Log-likelihood  | -78050           | -77546           | -77014          |
|                 | [-81289, -75091] | [-80815, -74345] | [-79999, -74172] |
| AIC             | 156103           | 155094           | 154034          |
|                 | [150185, 162580] | [148691, 161632] | [148351, 160005] |
| BIC             | 156113           | 155104           | 154065          |
|                 | [150195, 162590] | [148702, 161642] | [148381, 160035] |
This table reports model selection criteria for three optimized models. The numbers in parentheses are 95% confidence intervals, corrected for clustering by accounts and sell dates.

Figure A3. $P(Gain)$ and $P(Loss)$ as a function of portfolio composition for the one-stage, two-stage, and mixture models, and the empirical data. The one-stage, two-stage, and empirical columns repeat Figures 1, 2, and 5.
## Table A6. Regression Table for Four Logistic Models

<table>
<thead>
<tr>
<th>IV</th>
<th>Models</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-1.776</td>
<td>-2.806</td>
<td>-0.825</td>
<td>-2.969</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-1.847, -1.704]</td>
<td>[-2.889, -2.724]</td>
<td>[-0.909, -0.742]</td>
<td>[-3.058, -2.88]</td>
</tr>
<tr>
<td>Gain</td>
<td></td>
<td>0.593</td>
<td>0.042</td>
<td>1.041</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.522, 0.665]</td>
<td>[-0.079, 0.163]</td>
<td>[0.932, 1.15]</td>
<td>[0.011, 0.258]</td>
</tr>
<tr>
<td>√Holding days</td>
<td></td>
<td>-0.009</td>
<td>-0.006</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.013, -0.005]</td>
<td>[-0.009, -0.003]</td>
<td>[-0.008, -0.003]</td>
<td>[-0.008, -0.002]</td>
</tr>
<tr>
<td>Gain x Return</td>
<td></td>
<td>0.701</td>
<td>0.908</td>
<td>0.942</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.484, 0.918]</td>
<td>[0.685, 1.131]</td>
<td>[0.717, 1.167]</td>
<td>[0.739, 1.202]</td>
</tr>
<tr>
<td>Loss x Return</td>
<td></td>
<td>-0.095</td>
<td>-0.268</td>
<td>-0.334</td>
<td>-0.317</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.445, 0.254]</td>
<td>[-0.626, 0.091]</td>
<td>[-0.82, 0.153]</td>
<td>[-0.926, 0.091]</td>
</tr>
<tr>
<td>Gain x Return 20</td>
<td></td>
<td>1.419</td>
<td>1.126</td>
<td>1.158</td>
<td>1.129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.244, 1.594]</td>
<td>[0.964, 1.287]</td>
<td>[1.007, 1.309]</td>
<td>[0.969, 1.289]</td>
</tr>
<tr>
<td>Loss x Return 20</td>
<td></td>
<td>-0.331</td>
<td>-0.428</td>
<td>-0.456</td>
<td>-0.451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.552, -0.111]</td>
<td>[-0.648, -0.208]</td>
<td>[-0.685, -0.228]</td>
<td>[-0.677, -0.225]</td>
</tr>
<tr>
<td>Gain x Volatility 20 (× 1000)</td>
<td></td>
<td>0.056</td>
<td>0.110</td>
<td>0.106</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.048, 0.063]</td>
<td>[0.085, 0.131]</td>
<td>[0.090, 0.126]</td>
<td>[0.093, 0.129]</td>
</tr>
<tr>
<td>Loss x Volatility 20 (× 1000)</td>
<td></td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0, 0.001]</td>
<td>[0, 0.001]</td>
<td>[0, 0.001]</td>
<td>[0, 0.001]</td>
</tr>
<tr>
<td>Gain x Return × √Holding days</td>
<td></td>
<td>-0.021</td>
<td>-0.031</td>
<td>-0.031</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.029, -0.013]</td>
<td>[-0.039, -0.022]</td>
<td>[-0.039, -0.022]</td>
<td>[-0.041, -0.023]</td>
</tr>
<tr>
<td>Loss x Return × √Holding days</td>
<td></td>
<td>-0.009</td>
<td>0.011</td>
<td>0.012</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.025, 0.007]</td>
<td>[-0.006, 0.027]</td>
<td>[-0.008, 0.032]</td>
<td>[-0.006, 0.03]</td>
</tr>
<tr>
<td>( N_{g, 2} ) x Gain</td>
<td></td>
<td>7.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[7.074, 7.467]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{g, 2} ) x Loss</td>
<td></td>
<td>-0.339</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.357, -0.32]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{L} ) x Gain</td>
<td></td>
<td>-0.258</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.277, -0.238]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{L} ) x Loss</td>
<td></td>
<td>-0.041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.056, -0.026]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{g, 2} ) x Gain</td>
<td></td>
<td>-0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.066, -0.039]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{g, 2} ) x Loss</td>
<td></td>
<td>2.991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[2.916, 3.065]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{L} ) x Gain</td>
<td></td>
<td>1.664</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.581, 1.747]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N_{L} ) x Loss</td>
<td></td>
<td>0.622</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.548, 0.696]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LogLikelihood</td>
<td></td>
<td>-87423.83</td>
<td>-78074.19</td>
<td>-78282.05</td>
<td>-76923.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-76423.9, -78074.19]</td>
<td>[-78074.19, -78282.05]</td>
<td>[-78282.05, -76923.32]</td>
<td>[-76923.32, -76923.32]</td>
</tr>
<tr>
<td>R2</td>
<td></td>
<td>0.030</td>
<td>0.134</td>
<td>0.132</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.003, 0.034]</td>
<td>[0.114, 0.154]</td>
<td>[0.115, 0.151]</td>
<td>[0.143, 0.159]</td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>174869.67</td>
<td>156174.38</td>
<td>156594.10</td>
<td>153876.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[174169.57, 156514.38]</td>
<td>[155814.38, 156854.10]</td>
<td>[156314.10, 153536.64]</td>
<td>[153536.64, 153876.64]</td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td>174980.89</td>
<td>156305.82</td>
<td>156745.77</td>
<td>154028.31</td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td>181896</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A6 Ruling Out an Alternative Explanation on Realization Utility

Barberis and Xiong (2012) proposed a model of realization utility in which investors derive utility by realizing gains or losses from their portfolio. Here we consider whether an extension of the realization utility model may offer an alternative explanation of the within-domain sensitivity that we reported.

Suppose that investors who have many more stocks in loss than in gain in their portfolio choose to myopically realize one of the gains in order to compensate for the negative feelings associated with having many unrealized losses. It therefore follows that the disposition effect coefficient, $\beta$, may positively correlates with $N_L/N_G$. In other words, the larger the number of losses in the portfolio ($N_L$) relative to the number of gains ($N_G$), the larger the disposition effect on individual stock level ($\beta$). In this alternative model, $\beta$ is a function of $N_G$ and $N_L$. Also, assuming that $\beta$ is completely proportional to $N_L/N_G$, the disposition coefficient $\beta = (N_L/N_G) \times A$, where $A$ is a constant. Substituting this $\beta$ for the constant $\beta$ in the one-stage model (Equation 1):

$$(\text{Gain}) = \frac{1}{N_G + \frac{N_L}{\beta}}$$

we obtain $P(\text{Gain}) = \frac{1}{N_G(1+\frac{1}{A})}$. This result is identical to Equation 3 in the two-stage model that we proposed, where $B = A$. Therefore, if investors tend to compensate for their negative feelings caused by the presence of many losses in their portfolios by realizing a gain and if the likelihood of investors doing so is a function of $N_L/N_G$, the prediction from the myopic realization utility and that from the two-stage model is mathematically identical and both predict the inverse proportionality seen in Figure 5.

However, the assumption that $\beta$ is proportional to $N_L/N_G$ that underpins the alternative model may be too strong. That is, investors’ utility from unrealized gains and losses in a portfolio and the utility derived by realizing gains or losses should be, at least partially, dependent on the size of gains and losses. More specifically, if the alternative model is correct, we should expect that the larger the size of unrealized loss in the portfolio the larger the size of realized gain on the sell-day.
Figure A4 shows the size of realized gain/loss on the sell-day as a function of the size of unrealized gain/loss in the portfolio. The curve is flat for the portfolio-level losses and has an upward slope in the gain domain. This is not consistent with the prediction of the myopic realization utility model.

![Figure A4. The size of realized gain/loss on the sell-day as a function of the size of unrealized gain/loss in the portfolio. The blue line is a prediction form a local regression. The shaded areas are bootstrapped 95% confidence intervals.](image)

The x-axis variable in Figure A4 correlates with $N_G$ and $N_L$ and may be confounded with the disposition effect where a stock with a large unrealized gain was realized on the day, and thus, one large gain contributes to both the x- and y-axes. In order to account for this, we extracted portfolios with $N_G = 1$ and $N_L = 1$ and calculated $P(Gain)$ as a function of the size of unrealized gain/loss in the portfolio (i.e., the aggregation of a gain and a loss within a portfolio). If the alternative model is correct, we would expect that the larger the size of an unrealized loss in the portfolio the larger $P(Gain)$. 
Figure A5 plots the results, showing no association between portfolio-level unrealized gain/loss and the likelihood of a gain being realized. This is again not consistent with the prediction of the myopic realization utility model.

Figure A5. $P(Gain)$ as a function of the size of unrealized gain/loss in the portfolio. The blue line is a prediction form a local regression. The shaded areas are bootstrapped 95% confidence intervals.

Whereas it is technically possible that investors’ utility from unrealized and realized gain/loss is a function of $N_G$ and $N_L$, regardless of the size of unrealized and realized gain/loss, such a utility function seems unrealistic. On the other hand, as discussed in Section 5, given that an investor already decided to sell one of gains, it is trivial that $P(Gain)$ is, on average, inversely proportional to $N_G$. Thus we consider that the two-stage model explanation is more psychologically plausible than the alternative explanation based on the myopic realization utility.