Realization Utility

Nicholas Barberis and Wei Xiong

October 2007*

Abstract

We study the possibility that, aside from standard sources of utility, investors also derive utility from *realizing* gains and losses on individual investments that they own. We propose a tractable model of this "realization utility," derive its predictions, and show that it can shed light on a number of puzzling facts. These include the poor trading performance of individual investors, the disposition effect, the greater turnover in up markets, the negative premium to volatility in the cross-section, and the heavy trading of highly valued assets. Underlying some of these applications is one of our model's more novel predictions: that, even if the form of realization utility is linear or concave, investors can be risk-*seeking*.

*Comments are welcome at nick.barberis@yale.edu and wxiong@princeton.edu. We thank Bige Kahraman, Chris Rogers, and participants at the 2007 Cambridge-Princeton conference for helpful discussions.

1 Introduction

When economists model the trading behavior of individual investors, they typically assume that these investors derive utility only from consumption, or from total wealth. In this paper, we study the possibility that investors also derive utility from another source, namely from *realized gains and losses* on risky assets that they own. Suppose, for example, that an investor buys a stock, and then, a few months later, sells his position. We analyze a model in which the investor gets a jolt of utility right then, at the moment of sale, and where the utility term depends on the size of the gain or loss realized – utility is positive if the stock is sold at a gain relative to purchase price, and negative otherwise. We label this source of utility "realization utility."

There are several reasons why an investor might derive utility from realizing a gain or loss. If he sells a stock at a gain, he can tell himself that he is a savvy investor, raising his self-esteem. The sale also gives him a piece of news to boast about to family and friends. While an investor can certainly also feel good about a stock trading at a *paper* gain, these feelings are likely to be more pronounced at the moment of sale. When an asset is sold, there is a sense that the transaction is "complete," making it easier to claim credit for a successful investment.

Practitioners have long argued that realization utility plays an important role in individual decision making. For example, in a well-known manual for stock brokers, Gross (1982) discusses the pain associated with realizing a loss:

> *"Most clients, however, will never sell anything at a loss... Investors are reluctant to accept and realize losses because the very act of doing so proves that their first judgment was wrong... Investors who accept losses can no longer prattle to their loved ones, 'Honey, it's only a paper loss. Just wait. It will come back.' "*

Introspection and casual observation, then, suggest that realization utility may be a significant driver of trading behavior, and hence that it merits a more systematic analysis. In this paper, we do three things. First, we develop a tractable model of realization utility, one that is sophisticated enough to capture many features of actual trading, but also simple enough to allow for an analytical solution. Second, we lay out the model's predictions. And third, we link these predictions to a wide range of applications.

In its simplest version, our model is an infinite-horizon framework in which an investor switches back and forth between a stock and a risk-free asset. Whenever he liquidates his stock holdings, he receives a jolt of utility based on the size of the gain or loss realized, and pays a proportional transaction cost. In one extension of the basic model, we also allow for a random liquidity shock which forces the investor to sell any outstanding position in stock and to exit the stock market. In another extension, we also require the investor to park the proceeds of a stock sale in cash for a certain period of time before re-entering the stock market – a device intended to capture the idea that it can take time to dig up attractive stock market opportunities.

At each moment, the investor makes his allocation decision by maximizing the discounted sum of expected future realization utility flows. For much of the paper, we assume a *linear* functional form for realization utility, and standard exponential discounting.

We find that, in our model, the investor voluntarily sells his stock holdings only when the stock is trading at a sufficiently large gain relative to purchase price. We show how this "liquidation point" – the percentage gain in price, relative to purchase price, at which the investor is willing to sell – depends on the stock's expected return, its standard deviation, the investor's time discount rate, the level of transaction costs, and the frequency of liquidity shocks. Our model also allows us to compute the probability that, within any given interval after first buying a stock, the investor sells it. We show how this probability – a measure of trading frequency – depends on the aforementioned factors.

The model makes a number of interesting predictions. One of the more striking is that, even if realization utility has a linear functional form, the investor can be risk-*seeking*: all else equal, his time 0 value function is increasing in the volatility of the stock available for trading. The intuition is straightforward. A highly volatile stock offers the chance of a large gain, which the investor can enjoy realizing. Of course, it may also experience a large drop in value; but in that case, the investor will simply postpone selling the stock until he is forced to sell by a liquidity shock. Any realized loss therefore lies in the heavily discounted future and does not scare the investor very much. Overall, then, the investor prefers more volatility to less.

A related intuition underlies another of the model's predictions: that the investor is willing to buy a stock with a *negative* average excess return, so long as its volatility is sufficiently high. The model also predicts that more volatile stocks will be traded more frequently: roughly speaking, a more volatile stock reaches its liquidation point more rapidly.

We link our model to a wide range of financial phenomena. In particular, we argue that it offers a way of thinking about the subpar trading performance of individual investors, the disposition effect, the greater turnover in bull markets than in bear markets, the negative volatility premium documented by Ang et al. (2005), and the heavy trading associated with highly valued assets – as, for example, in the technology sector in the late 1990s.

To understand this last application, note that, in an economy where many investors care about realization utility, more volatile stocks will be both more heavily traded – such stocks reach their liquidation points faster – *and* more highly valued: since realization utility investors like volatility, they will collectively push the prices of volatile stocks up. Our model therefore predicts a coincidence of high valuations and heavy trading; and moreover, that this phenomenon will occur for assets whose fundamentals are particularly uncertain. Under this view, the late 1990s were years where realization utility investors, attracted by the high uncertainty of technology stocks, bought these stocks, pushing their prices up; as (some of) these stocks rapidly reached their liquidation points, the realization utility investors sold them, and then immediately bought new ones.

Although we work mainly with exponential time discounting, we also consider the case of hyperbolic time discounting. While hyperbolic discounting has been linked to a number of economic phenomena, researchers have not, as yet, found many applications for it within the context of finance. We show that, as soon as we allow for realization utility, hyperbolic discounting can have significant effects.

As noted above, practitioners have for decades noted the potential importance of realization utility. In the academic literature, an early discussion of this idea can be found in Shefrin and Statman (1985). They propose it, in combination with prospect theory, as a way of understanding the disposition effect, and present a two-period numerical example. More recently, Barberis and Xiong (2006) analyze a two-period model of realization utility, again in combination with prospect theory, and again with the disposition effect as the eventual application.

In this paper, we offer the first comprehensive analysis of realization utility. We move beyond the two-period setting and work in an infinite horizon framework. We allow for transaction costs, for random liquidity shocks, and for a time interval during which the investor searches for attractive stock market opportunities. We analyze the investor's trading strategy along several dimensions, including trading frequency. We consider a wide range of applications, of which the disposition effect is just one. And while our framework allows for a prospect theory functional form, we work mainly with a linear functional form, and investigate what it, in combination with a positive time discount factor, can predict.

2 A Model of Realization Utility

We now present a model of trading behavior in which the investor cares about realization utility. In Section 2.1, we lay out the basic version of the model. In Section 2.2, we generalize the model along two dimensions. We adopt this "incremental" strategy of starting with the simplest model so as to illustrate the effect of each additional feature as clearly as possible.

We make two fundamental modeling assumptions. First, we assume that the carriers of realization utility are *gains and losses* measured relative to purchase price, rather than absolute wealth levels. If realization utility is the idea that the investor derives pleasure from completing a successful investment in some asset, it is natural that how good he feels at the moment of sale is a function of the asset's change in value since purchase.

Second, we assume that realization utility is defined at the level of an individual asset. Again, if realization utility is the idea that an investor feels good when he completes a successful investment in some asset, it is natural that utility is defined at the asset level, even if the asset is just one of many in his portfolio. The idea that an agent might get utility from the outcome of one specific asset that he owns is sometimes known as "narrow framing." In short, then, realization utility leads naturally to narrow framing.

Taken together, these assumptions mean that the utility specification in our model differs from more traditional specifications in three ways: in that utility is defined over gains and losses rather than wealth levels; in that utility is defined at the level of individual assets; and in that the utility specification makes a distinction between realized and paper gain/losses, and defines utility only over *realized* gains and losses.

Another modeling choice concerns the functional form for realization utility. Since it is unclear what this functional form should be, we focus on the simplest possibility, a linear functional form, and show that, even under this assumption, realization utility has a range of novel implications. Later in the paper, we consider some alternative specifications. In particular, while Section 2 adopts the linear functional form throughout, Section 3.1 considers a piecewise linear specification. In Section 3.2, we vary another dimension of preferences, and replace exponential time discounting with hyperbolic time discounting.

Our model is related to Kahneman and Tversky's (1979) prospect theory, but only weakly so. The common feature is that utility is defined over gains and losses, rather than absolute wealth levels, but the similarities end there. By focusing on a linear functional form for utility, we largely ignore key elements of prospect theory such as loss aversion and diminishing sensitivity to gains and losses. Even our assumption that realization utility is defined over gains and losses does not have to be motivated from prospect theory – by thinking about what realization utility is trying to capture, it quickly becomes clear that it is best defined over gains and losses.

2.1 The basic model

We use a continuous-time framework because this allows us to solve the model analytically. We have also studied the discrete-time analog of our model. The results are similar, but can only be obtained numerically.

Consider an investor who starts at time 0 with wealth W_0. At each time $t \geq 0$, he has two investment options: a risk-free asset, which offers a net return of zero; and a risky asset – a stock, say – whose price S_t follows

$$\frac{dS_t}{S_t} = \mu dt + \sigma dZ_t. \tag{1}$$

For simplicity, we assume that, at each time t, the investor either allocates all of his wealth to the risk-free asset or all of his wealth to the stock: no intermediate allocations are allowed. We also suppose that, if the investor sells his position in the stock at time t, he pays a proportional transaction cost, kW_t, $0 \leq k < 1$, where W_t is time t wealth. The investor's wealth therefore evolves according to

$$\frac{dW_t}{W_t} = \theta_t(\mu dt + \sigma dZ_t) - kI_{\{l_t=1\}}, \tag{2}$$

where θ_t takes the value 1 if he is holding the stock at time t, and 0 otherwise; and where l_t takes the value 1 if he sells the stock at time t, and 0 otherwise.

An important variable in our model is B_t. This variable, which is defined only if the investor is holding stock at time t, measures the cost basis of the stock position, in other words, the amount of money the investor put into the time t stock position at the time he bought it. Formally, if $\theta_t = 1$,

$$B_t = W_s, \text{ where } s = \max\{\tau \epsilon [0, t) : \theta_\tau = 0\}. \tag{3}$$

The key feature of our model is that the investor derives utility from realizing a gain or loss. Specifically, whenever he switches his wealth from the stock into cash, he receives a burst of utility given by

$$u((1 - k)W_t - B_t). \tag{4}$$

The argument of the utility term is the size of the realized gain or loss: the investor's wealth at the moment of sale, after the transaction cost, $(1-k)W_t$, minus the cost basis of the stock investment B_t.[1]

Suppose that, at time t, the investor's wealth is allocated to the stock. The investor's value function is a function of the current asset value, W_t, and of the asset's cost basis, B_t. We denote the value function as $V(W_t, B_t)$.

We further assume

$$V(W, W) > 0. \tag{5}$$

[1]The investor only receives realization utility when he liquidates a position in stock and puts the proceeds into the risk-free asset, not when he sells the risk-free asset and puts the proceeds into stock. The reason is that, since the risk-free rate is zero, the realized gain or loss from selling the risk-free asset is always zero. We also assume that the investor does not incur a transaction cost from selling the risk-free asset.

Note that $V(W, W)$ is the value function from investing in the stock *now*, so that the asset's current value and cost basis are both equal to current wealth W. Given a positive time discount rate, condition (5) implies two things. First, it implies that, at time 0, the investor allocates his wealth to the stock: since the risk-free asset generates no utility flows, he allocates to the stock as early as possible, in other words, at time 0. Second, condition (5) implies that, if, at any time $t > 0$, the investor sells a position in stock, he will then immediately re-establish the position. We verify the validity of condition (5) later.

We can now formulate the investor's decision problem. At time t, the investor solves

$$V(W_t, B_t) = \max_{\tau \geq t} E_t \left\{ e^{-\delta(\tau - t)} \left[u((1-k)W_\tau - B_\tau) + V((1-k)W_\tau, (1-k)W_\tau) \right] \right\}, \tag{6}$$

subject to (2), (3), and

$$\begin{aligned} \theta_s &= 1, t \leq s < \tau \\ \theta_\tau &= 0. \end{aligned} \tag{7}$$

In words, at time t, the investor chooses the optimal time τ, a random time in the future, at which to realize the gain or loss in his current position. When he liquidates his position at time τ, he receives a burst of utility $u((1-k)W_\tau - B_\tau)$ and a cash balance of $(1-k)W_\tau$ which he immediately reinvests in the stock. The parameter δ is the time discount rate. To ensure that the investor does not hold his time 0 stock position forever, without selling it, we assume

$$\delta > \mu. \tag{8}$$

We consider two functional forms for the utility term $u(\cdot)$. Throughout Section 2, we consider the linear case

$$u(x) = x. \tag{9}$$

As noted earlier, we focus on this case because it is the simplest one, and because we want to show that we do not need strong assumptions about $u(\cdot)$ in order to derive interesting results. In Section 3, however, we also consider a piecewise linear specification.

At this point, we note an important caveat. For the sake of tractability, the investor in our model derives utility *only* from realized gains and losses: other, more standard sources of utility are ignored. As such, our framework is not suitable for thinking about how an investor would allocate his *overall* wealth: rather, it is designed to shed light on how investors trade stocks in their brokerage accounts. Put differently, we should not think of W_0 as the investor's total wealth, but as that portion of his total wealth that he allocates to a brokerage account.

So far, we have interpreted the mathematical structure in (2), (3), and (6) in terms of a *one*-stock model, so that, over time, the investor switches in and out of the *same* stock. However, the same mathematical structure also admits an alternative interpretation, which we prefer, and which we adopt going forward. Under this interpretation, the economy contains *many* stocks, all of which have the same return distribution in (1). At any time $t \geq 0$, the investor can either have all of his wealth in the risk-free asset, or all of his wealth in one of the stocks. And whenever the investor moves his wealth into a stock, the stock is a *new* stock, one that he has not previously owned.

Why is this a better interpretation? Under the one-stock interpretation, the objective function in (6) says that, if the investor liquidates his position in a stock, and then, a minute later, buys back the same stock, he nonetheless derives utility from the gain or loss realized at the moment of sale. This seems psychologically implausible. It is hard to imagine that the investor can derive realization utility from the sale of a stock if he then immediately buys the stock back: he can hardly claim credit for a successful "completed" transaction if he then immediately reopens the transaction. If, however, he sells a stock and then, a minute later, buys a *new* stock, it is more reasonable that he would derive utility from the gain or loss realized at the moment of sale: since the new stock is a different one, it is easier to think of the sale of the previous stock as a completed transaction.

While the multi-stock interpretation seems to us the more plausible one, it does make an implicit assumption, namely that, whenever the investor sells a stock, there is immediately available, for potential purchase, another stock with the same return characteristics. We maintain this assumption for now, but relax it in Section 2.2.[2]

The proposition below summarizes the solution to the decision problem in (6). The variable

$$g_t = \frac{W_t}{B_t} \tag{10}$$

– in words, the percentage change in value, since purchase, of the risky asset the investor is holding at time t – plays an important role in the solution.

Proposition 1: An investor with the decision problem in (6) will sell a position in stock once the gain $g_t = W_t/B_t$ reaches a liquidation point $g_* > 1$, where g_* is the unique solution to

$$\frac{1-k}{\gamma_1 - 1} g_*^{-(\gamma_1 - 1)} + (1-k)\, g_* - \frac{\gamma_1}{\gamma_1 - 1} = 0, \tag{11}$$

[2]Even under the multi-stock interpretation, we still assume, for simplicity, that the investor holds at most one stock at any time. Under some conditions, however, the solution to the problem in (6) also determines how an investor would trade in a setting where he holds *several* stocks concurrently. Suppose that the investor starts with wealth of mW_0, and spreads this wealth across m stocks, investing W_0 in each one. Now suppose that each of these m packets of wealth is managed as described earlier in the section: if the investor sells one of the stocks, he immediately reinvests the proceeds in another stock. Moreover, he derives utility separately from the realized gain or loss on each individual stock. The solution to (6) then also describes how the investor trades each of his stocks in this multiple-concurrent-stock setting.

where

$$\gamma_1 = \frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2\delta\sigma^2} - \left(\mu - \frac{1}{2}\sigma^2\right)\right] > 0. \tag{12}$$

The investor's value function is $V(W_t, B_t) = B_t U(g_t)$, where

$$U(g_t) = \begin{cases} \frac{(1-k)}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)} g_t^{\gamma_1} & \text{if} \quad g_t < g_* \\ (1-k)g_t(1 + U(1)) - 1 & \text{if} \quad g_t \geq g_* \end{cases}. \tag{13}$$

Furthermore, the investor will only invest in stock at time 0 if

$$U(1) = \frac{1-k}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)} > 0. \tag{14}$$

We prove the proposition in the Appendix. In brief, the proof proceeds by conjecturing that the investor sells his stock position once g_t exceeds some $g_* > 0$; by constructing the value function, first for the region below g_*, and then for the region above g_*; by requiring that the value function is continuous and continuously differentiable at g_*; and finally, by verifying that the constructed value function is indeed optimal.

Results

The shaded area in the top-left graph in Figure 1 shows the range of values of the stock's expected return μ and standard deviation σ for which $U(1) > 0$. In words, this is the range for which the investor is willing both to buy stock at time 0 *and* to sell it at some finite liquidation point. To do this calculation, we need to assign values to the two remaining parameters, δ and k. We set the time discount factor to $\delta = 0.08$, and the transaction cost to $k = 0.01$, which is of a similar order of magnitude to the transaction cost estimated by Barber and Odean (2000) for discount brokerage customers.

The graph illustrates an important feature of our model, namely that the investor is willing to invest in a stock even if it has a *negative* expected return. The intuition is simple. So long as the stock's standard deviation σ is positive, even a negative expected return stock has *some* chance of reaching the liquidation point g_*, at which time the investor can enjoy realizing the gain. Of course, more likely than not, the stock will lose value. However, since the investor does not realize losses, this will never bring him any disutility. Overall, then, investing in stock, even if it has a negative expected return, is a better option than investing in the risk-free asset, which offers zero utility for sure. In later sections, we show that this intuition continues to hold in more general models. For example, it holds even in the presence of liquidity shocks which force the investor to sell his holdings, whatever their value.

When $k = 0$, the condition that $U(1)$ exceed 0 reduces to the condition $\gamma_1 > 1$, which, in turn, reduces to $\mu < \delta$. Since the transaction cost k used in Figure 1 is low, we would expect $\mu < \delta$ to approximate the shaded region in the top-left graph quite well. Visually, this is the case. Note also that in the unshaded area to the right of the shaded area, the investor buys stock at time 0 but then holds it forever.

Figure 2 shows how the liquidation point g_* and time 0 utility $U(1)$ depend on the model parameters. The graphs on the left correspond to the liquidation point, and those on the right, to time 0 utility. For now, we focus on the solid lines; we discuss the dotted lines in Section 2.2.

To construct the graphs, we start with a set of benchmark parameters. We use the same benchmark parameters throughout the paper. We set the average excess return on stock to $\mu = 0.03$ and its standard deviation to $\sigma = 0.5$. We use a time discount factor of $\delta = 0.08$ and a transaction cost of $k = 0.01$. The graphs in Figure 2 show what happens as we vary each of μ, σ, and δ in turn, keeping the other parameters fixed at their benchmark levels.

The top graphs in Figure 2 show that, as we would expect, time 0 utility is increasing in the mean stock return μ. The liquidation point is also increasing in μ: if a stock offers a high average return and is trading at a gain, the investor is tempted to hold on to it, rather than sell it and incur a transaction cost.

The middle graphs illustrate one of the important predictions of our model: that, as stock return volatility goes up, the investor's time 0 utility also goes up. Put differently, even though the form of realization utility is linear, the investor is risk-*seeking*. While this is initially surprising, there is a simple intuition for it: a highly volatile stock offers the chance of a significant gain, which the investor can enjoy realizing. Of course, it also offers the chance of a significant loss. But the investor does not realize losses, and so will never experience any disutility. More volatile stocks are therefore more attractive. We show later that this result continues to hold in more general models; for example, it holds even in the presence of random liquidity shocks. Note also that, from a mathematical perspective, this prediction is a consequence of the fact that, while instantaneous utility is linear, the value function in (13) is convex.

The bottom graphs show that, when the investor discounts the future more heavily, utility falls, as does the liquidation point. An investor with a higher discount rate is more impatient, and therefore cannot wait as long before realizing a gain.

The top graphs in Figure 3 show how the liquidation point and initial utility depend on the transaction cost k. As expected, a higher transaction cost lowers the investor's time 0 utility. It also increases the liquidation point: given that it is costly to liquidate a position, the investor waits longer before doing so.

It is useful to note what happens when the transaction cost falls to zero. The top-left graph in Figure 3 suggests that, in this case, the liquidation point falls to $g_* = 1$. It is simple to confirm that, when $k = 0$, equation (11) is indeed satisfied by $g_* = 1$. In this case, then, the investor realizes all gains immediately. Also, as noted earlier, in this case, the investor is willing to invest in stock at time 0 so long as $\mu < \delta$.

2.2 A more general model

We now make our model more realistic by introducing some additional features of trading. First, we introduce a "search interval." The model of Section 2.1 assumes that, whenever the investor sells a stock, he can immediately reinvest the proceeds in another stock with the same return distribution. In reality, it may take the investor some time to find another stock with a similarly attractive return distribution. We therefore extend the basic model so that, whenever the investor sells a stock, he then has to park the proceeds in the risk-free asset for an interval of length T, during which time he searches for a new stock to invest in. Only at the end of the interval does he invest in the new stock.

In the model of Section 2.1, the wealth in the investor's brokerage account affects his utility only through the channel of realized gains and losses. We do not model the idea that, since it forms part of his total wealth, the money in the brokerage account may affect the investor's consumption, which, in turn, is a significant source of utility for him.

To capture, in reduced form, the idea that the investor might need the money in his brokerage account for consumption purposes, we now also allow for the possibility that the investor receives an exogeneous liquidity shock at some time τ' which forces him to liquidate all of his holdings and to exit the stock market. We assume that the liquidity shock arrives according to a Poisson process with parameter ρ.

With these two new features, the investor's decision problem becomes

$$\begin{aligned} V(W_t, B_t) &= \max_{\tau \geq t} \; E_t \Big\{ e^{-\delta(\tau - t)} \Big[u((1-k)W_\tau - B_\tau) I_{\{\tau < \tau'\}} \\ &\qquad + e^{-\delta T} V\left((1-k)W_\tau, (1-k)W_\tau\right) I_{\{\tau + T < \tau'\}} \\ &\qquad + u((1-k)W_{\tau'} - B_{\tau'}) I_{\{\tau \geq \tau'\}} \Big] \Big\} . \qquad (15) \end{aligned}$$

To understand this, suppose that the investor sells the stock early enough so that $\tau + T < \tau'$, in other words, so that the liquidity shock arrives not only after the sale, but also after the search interval that follows. In this case, only the first two of the three terms within the square parentheses are non-zero: the investor receives utility $u((1-k)W_\tau - B_\tau)$ at the moment of sale, as well as $e^{-\delta T} V\left((1-k)W_\tau, (1-k)W_\tau\right)$, the value function at the end of the search interval, discounted back.[3]

[3] If the risk-free rate were non-zero, the $V\left((1-k)W_\tau, (1-k)W_\tau\right)$ term would also need to reflect com-

If the investor sells the stock later, so that $\tau < \tau' < \tau + T$, only the first of the three terms within the square parentheses is non-zero: the investor receives realization utility at the moment of sale, but, since the liquidity shock arrives during the search interval, he receives nothing else. If $\tau > \tau'$, the investor is forced out of the stock market by a liquidity shock and receives $u((1-k)W_{\tau'} - B_{\tau'})$ from the gain or loss at the moment of exit.

In the Appendix, we prove:

Proposition 2: Unless forced to exit the stock market by a liquidity shock, an investor with the decision problem in (15) will sell a position in stock once the gain $g_t = W_t/B_t$ reaches a liquidation point $g_* > 1$. The value function is $V(W_t, B_t) = B_t U(g_t)$, where

$$U(g_t) = \begin{cases} c_1 g_t^{\gamma_1} + \frac{\rho(1-k)}{\rho+\delta-\mu} g_t - \frac{\rho}{\rho+\delta} & \text{if} \quad g_t < g_* \\ (1-k) g_t (1 + e^{-(\rho+\delta)T} U(1)) - 1 & \text{if} \quad g_t \geq g_* \end{cases}, \tag{16}$$

where

$$\gamma_1 = \frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2(\rho+\delta)\sigma^2} - \left(\mu - \frac{1}{2}\sigma^2\right)\right] > 0, \tag{17}$$

and where c_1 and g_* are given by

$$c_1 = \frac{(1-k)\left[1 + e^{-(\rho+\delta)T}\left(\frac{\rho(1-k)}{\delta+\rho-\mu} - \frac{\rho}{\rho+\delta}\right)\right] - \frac{\rho(1-k)}{\delta+\rho-\mu}}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)e^{-(\rho+\delta)T}} \tag{18}$$

$$0 = c_1 g_*^{\gamma_1}(1-\gamma_1) - \frac{\rho}{\rho+\delta} + 1. \tag{19}$$

Results

To illustrate the effect of the two new features of our model – the search interval and the liquidity shock – we introduce them one at a time. We start by setting $T = 0.1$ and $\rho = 0$, so that there is a search interval but no liquidity shock. A value of $T = 0.1$ means that the investor has to wait just over a month after selling one stock before he is able to buy another.

The shaded area in the top-right graph in Figure 1 plots the range of values of μ and σ for which $U(1) > 0$, in other words, the range for which the investor is willing both to buy stock at time 0 and to sell it at a finite liquidation point. In these calculations, we set the two remaining parameters, δ and k, to the benchmark values from before, 0.08 and 0.01, respectively.

pounded interest.

The range of values of μ and σ for which $U(1) > 0$ is very similar to that in the model of Section 2.1. It is straightforward to show that, when $k = 0$ and $T = 0$, the condition $U(1) > 0$ reduces to $\mu < \delta$. Since, in the top-right graph of Figure 1, both k and T are low, we would expect the graph to be well approximated by $\mu < \delta$. Visually, this is the case.

The middle two graphs in Figure 3 show how the liquidation point g_* and initial utility $U(1)$ depend on the length of the search interval T. In producing these graphs, the remaining parameters are set to the benchmark levels noted earlier, namely

$$(\mu, \sigma, \delta, k) = (0.03, 0.5, 0.08, 0.01). \tag{20}$$

The middle-left graph shows that the liquidation point rises with T: the longer the investor has to wait, in cash, after selling a position in stock, the more reluctant he will be to sell the stock in the first place. This result can also be seen analytically: when $\rho = 0$, the liquidation point g_* from equations (18) and (19) satisfies

$$\frac{\partial g_*}{\partial T} = \frac{\frac{\delta e^{-\delta T}}{\gamma_1 - 1} g_*^{-(\gamma_1 - 1)}}{1 - e^{-\delta \tau} g_*^{-\gamma_1}} > 0, \tag{21}$$

so that the liquidation point does indeed increase with the length of the search interval T. The middle-right graph shows that the investor's initial utility falls as T rises: the longer the interval over which the investor is barred from holding the attractive stock, the less happy he is.

The dotted lines in Figure 2 show how the liquidation point g_* and initial utility $U(1)$ vary with μ, σ, and δ in the presence of a search interval $T > 0$. In these calculations, we vary each of μ, σ, and δ in turn, keeping the other parameters fixed at their benchmark values, namely

$$(\mu, \sigma, \delta, k, T) = (0.03, 0.5, 0.08, 0.01, 0.1). \tag{22}$$

The dotted lines in the graphs on the left side of Figure 2 show that introducing a search interval preserves the basic relationship between g_* on the one hand and μ, σ, and δ on the other; the effect is simply to shift the liquidation point up. It also preserves the basic relationship between $U(1)$ and μ, σ, and δ; it simply shifts the utility level down somewhat.

We now allow for both a search interval *and* a liquidity shock. Specifically, we maintain $T = 0.1$ and assign ρ a benchmark value of 0.1, which implies that the probability of a liquidity shock over the course of a year is $1 - e^{-0.1} \approx 0.1$.

The shaded area in the middle-left graph in Figure 1 shows the range of values of μ and σ for which $U(1) > 0$. We set the remaining parameters, δ and k, to their benchmark values of 0.08 and 0.01, respectively. The figure shows that, relative to the case where there is no liquidity shock, the investor is now only willing to invest in a negative expected return stock if its standard deviation is sufficiently high: in the presence of a liquidity shock, a negative expected return stock is less attractive because it raises the chance that the investor will be forced to make a painful exit from a losing position. Only if the stock has a high standard deviation, so that it also offers a chance of a sizeable gain which the investor can enjoy realizing, will he invest in it.

The bottom graphs in Figure 3 show how the liquidation point g_* and initial utility $U(1)$ depend on the intensity of the liquidity shock ρ. Here, the remaining parameters are set to the benchmark values in (22).

The bottom-left graph shows that the liquidation point depends on ρ in a non-monotonic way. There are two factors at work here. As the liquidity shock intensity ρ goes up, the liquidation point initially falls. One reason the investor delays realizing a gain is the transaction cost that a sale entails. In the presence of liquidity shocks, however, the investor knows that he is likely to be forced out of the stock market at some point. The present value of the transaction costs he expects to pay is therefore lower than in the absence of liquidity shocks. He is therefore willing to realize gains sooner.

At higher levels of ρ, however, there is a second factor which makes the investor more patient: if he is holding a stock with a gain, he is reluctant to exit the position, because he will then have to reinvest in another stock, which might do poorly, and from which he might be forced to exit at a loss by a liquidity shock. This factor pushes the liquidation point back up.

The bottom-right graph shows that, as the liquidity shock intensity ρ rises, the agent's utility falls. Since a liquidity shock may force the investor to exit the stock market with a painful loss, it lowers his utility.

The solid lines in Figure 4 show how the liquidation point g_* and initial utility $U(1)$ depend on the parameters μ, σ, and δ in the presence not only of transaction costs and a search interval, but also a liquidity shock, so that $\rho > 0$. In these calculations, we vary each of μ, σ, and δ in turn, keeping the remaining parameters fixed at the benchmark values

$$(\mu, \sigma, \delta, k, T, \rho) = (0.03, 0.5, 0.08, 0.01, 0.1, 0.1). \tag{23}$$

Comparing the solid lines in Figure 4 to the solid and dotted lines in Figure 2, we see that, with one exception, the possibility of a liquidity shock preserves the same basic relationship between g_* and $U(1)$ on the one hand, and μ, σ, and δ on the other. In particular, even in the presence of liquidity shocks, the investor's initial utility $U(1)$ is still strongly increasing in the stock's volatility σ. A liquidity shock may force the holder of a volatile stock to realize a large loss, and this would be painful; but it may also force him to realize a large gain, and this would be pleasurable. Given that utility is linear, liquidity shocks do not reverse the earlier relationship between initial utility and stock volatility.

The one difference between Figures 2 and 4 is in the way the investor's initial utility depends on the discount rate δ. In the absence of a liquidity shock – the case of Figure 2 – the investor's utility falls as he becomes more impatient, while in the presence of a liquidity shock, it rises.

3 Other Preference Specifications

In Section 2, we took the functional form for realization utility $u(\cdot)$ to be *linear*, and assumed exponential time discounting. We adopted the linear specification because of its simplicity and in order to show that we can derive interesting results without making strong assumptions about the form of $u(\cdot)$. In Section 3.1, we consider an alternative specification – piecewise linear utility – and show how it affects the results. In Section 3.2, we alter another dimension of preferences by replacing exponential discounting with hyperbolic discounting.

3.1 Piecewise linear utility

We start with the most complete model that we have developed so far – the model of Section 2.2, which allows for a transaction cost, a search interval, as well as a liquidity shock – and investigate what happens when $u(\cdot)$ has a piecewise linear form, rather than a linear one, so that the investor is "loss averse," or more sensitive to losses than to gains:

$$u(x) = \begin{cases} x & \text{if } x \geq 0 \\ \lambda x & \text{if } x < 0 \end{cases}, \qquad \lambda > 1. \tag{24}$$

The investor's decision problem is:

$$\begin{aligned} V(W_t, B_t) \;=\; \max_{\tau \geq t} \;\; E_t \Big\{ e^{-\delta(\tau - t)} \Big[& u((1-k)W_\tau - B_\tau) I_{\{\tau < \tau'\}} \\ & + e^{-\delta T} V\left((1-k)W_\tau, (1-k)W_\tau\right) I_{\{\tau + T < \tau'\}} \\ & + u((1-k)W_{\tau'} - B_{\tau'}) I_{\{\tau \geq \tau'\}} \Big] \Big\}. \end{aligned} \tag{25}$$

This is the same as equation (15) in Section 2.2, except that $u(\cdot)$ is no longer linear, but instead takes the form in (24).

In the Appendix, we prove:

Proposition 3: Unless forced to exit the stock market by a liquidity shock, an investor with the decision problem in (25) will sell a position in stock once the gain $g_t = W_t/B_t$ reaches a liquidation point $g_* > 1$. His value function is $V(W_t, B_t) = B_t U(g_t)$, where

$$U(g_t) = \begin{cases} c_1 g_t^{\gamma_1} + \frac{\rho\lambda(1-k)}{\rho+\delta-\mu} g_t - \frac{\rho\lambda}{\rho+\delta} & \text{if} \quad g_t \;\epsilon\; (0, \frac{1}{1-k}) \\ b_1 g_t^{\gamma_1} + b_2 g_t^{\gamma_2} + \frac{\rho(1-k)}{\rho+\delta-\mu} g_t - \frac{\rho}{\rho+\delta} & \text{if} \quad g_t \;\epsilon\; (\frac{1}{1-k}, g_*) \\ (1-k) g_t (1 + e^{-(\rho+\delta)T} U(1)) - 1 & \text{if} \quad g_t \;\epsilon\; (g_*, \infty), \end{cases} \tag{26}$$

where γ_1 is defined in equation (17), where

$$\gamma_2 = -\frac{1}{\sigma^2} \left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2(\rho+\delta)\sigma^2} + \left(\mu - \frac{1}{2}\sigma^2\right) \right] < 0, \tag{27}$$

and where c_1, b_1, b_2, and g_* are determined from

$$b_2 = \frac{1}{(\gamma_1 - \gamma_2)} \frac{(\lambda - 1)\rho(1-k)^{\gamma_2}(\mu\gamma_1 - \rho - \delta)}{(\rho+\delta-\mu)(\rho+\delta)} \tag{28}$$

$$(\gamma_1 - 1) b_1 g_*^{\gamma_1} + (\gamma_2 - 1) b_2 g_*^{\gamma_2} = \frac{\delta}{\rho+\delta} \tag{29}$$

$$b_1 \left(\frac{1}{1-k}\right)^{\gamma_1} + b_2 \left(\frac{1}{1-k}\right)^{\gamma_2} = c_1 \left(\frac{1}{1-k}\right)^{\gamma_1} + \frac{(\lambda-1)\mu\rho}{(\rho+\delta-\mu)(\rho+\delta)} \tag{30}$$

$$b_1 g_*^{\gamma_1} + b_2 g_*^{\gamma_2} + (1-k) g_* \frac{\mu - \delta}{\rho+\delta-\mu} + \frac{\delta}{\delta+\rho} = (1-k) g_* e^{-(\rho+\delta)T} \left(c_1 + \frac{\rho\lambda(\mu - k\rho - k\delta)}{(\rho+\delta)(\rho+\delta-\mu)} \right) \tag{31}$$

Specifically, given values for μ, σ, δ, k, T, ρ, and λ, we first use equation (28) to find b_2; we then obtain b_1 from equation (29); we then use equation (30) to find c_1; finally, equation (31) allows us to solve for the liquidation point g_*.

Results

The shaded area in the middle-right graph in Figure 1 shows the range of values of μ and σ for which $U(1)$, from (26), is positive. We set δ, k, T, and ρ to the benchmark values from before, namely 0.08, 0.01, 0.1, and 0.1, respectively. We further assign λ the benchmark value of 1.5. Relative to the middle-left graph – the graph for the model in Section 2.2, with liquidity shocks but no loss aversion – we see that the investor is now more reluctant to invest in stocks with negative expected returns. In the presence of liquidity shocks, a negative expected return stock is less attractive because it raises the chance that the investor will be forced to make a painful exit from a losing position. Loss aversion makes this prospect all the more unappealing. The investor will only invest in a negative expected return stock if it is highly volatile, so that it at least offers a chance of a sizeable gain which he can enjoy realizing.

The graphs in Figure 5 show how the liquidation point g_* and initial utility $U(1)$ depend on the degree of loss aversion λ. The dotted line corresponds to the case where there is no liquidity shock, $\rho = 0$, and the solid line to the case where there *is* a liquidity shock, $\rho > 0$. Specifically, for the dotted lines, we vary λ while maintaining

$$(\mu, \sigma, \delta, k, T, \rho) = (0.03, 0.5, 0.08, 0.01, 0.1, 0). \tag{32}$$

For the solid lines, we vary λ while maintaining

$$(\mu, \sigma, \delta, k, T, \rho) = (0.03, 0.5, 0.08, 0.01, 0.1, 0.1). \tag{33}$$

The dotted lines show that, in the absence of a liquidity shock, the degree of loss aversion λ is irrelevant: it has no effect at all on the liquidation point or on initial utility. The reason is simple: in the absence of a liquidity shock, the investor never realizes losses. How sensitive he is to losses is therefore irrelevant.

Loss aversion becomes relevant when $\rho > 0$. In the top-left graph, we see that, the more loss averse the investor is, the higher the liquidation point. The intuition is that, if the investor is holding a stock with a gain, he is reluctant to realize that gain, because if he does, he will have to reinvest in a new stock, which might go down, and from which he might be forced to exit at a loss by a liquidity shock.

The top-right graph shows that, as the degree of loss aversion goes up, the investor's utility falls: a high λ raises the possibility that the investor may be forced, by a liquidity shock, to make a painful exit from a losing position.

The dotted lines in Figure 4 show how the liquidation point g_* and initial utility $U(1)$ depend on μ, σ, and δ when the investor is loss averse. Here, we vary each of μ , σ, and δ in turn, keeping the other parameters fixed at their benchmark values,

$$(\mu, \sigma, \delta, k, T, \rho, \lambda) = (0.03, 0.5, 0.08, 0.01, 0.1, 0.1, 1.5). \tag{34}$$

Recall that the calculations for the solid lines in Figure 4 differ from those for the dotted lines only in the degree of loss aversion assumed: the solid lines correspond to linear $u(\cdot)$, so that $\lambda = 1$, while the dotted lines assume $\lambda = 1.5$. The dotted lines show that, for these benchmark parameters, allowing for loss aversion preserves the same basic relationship between g_* and $U(1)$ on the one hand, and μ, σ, and δ on the other. As expected from the graphs in Figure 5, increasing λ increases the liquidation point g_* and lowers utility $U(1)$.

The dotted line in the middle-right graph of Figure 4 shows that, for the benchmark values in (34), the investor's initial utility $U(1)$ is still increasing in stock volatility σ. Put differently, even though the functional form of realization utility is concave, the investor is risk-seeking. If the degree of loss aversion λ or the intensity of liquidity shocks ρ rises significantly, however, this relationship will reverse, so that $U(1)$ becomes a decreasing function of σ. A liquidity shock may force the holder of a volatile stock to realize a large gain, and this would be pleasurable; but it may also force to him to realize a large loss, and if λ rises significantly, this prospect will eventually make volatility unattractive.

3.2 Hyperbolic discounting

In the models we have presented so far, we have assumed *exponential* time discounting. This is a standard assumption in economic models, and one that implies time-consistent behavior. Recently, however, there has been mounting evidence that individual time preferences are better captured by *hyperbolic* time discounting. Relative to the exponential case, hyperbolic discounting places more weight on the present, as opposed to the future: under hyperbolic discounting, immediate rewards are especially attractive, and immediate costs, especially repellent.

While hyperbolic discounting has been linked to a number of economic phenomena, researchers have not, as yet, found many applications for it within the specific context of finance. We now show that, as soon as we allow for realization utility, hyperbolic discounting can play a significant role. Intuitively, hyperbolic discounting leads an investor who cares about realization utility to realize gains *earlier* than suggested by exponential discounting: realizing a gain *now* provides an immediate reward, and, under hyperbolic discounting, this is highly valued.

Harris and Laibson (2004) and Grenadier and Wang (2007) show how hyperbolic discounting can be introduced into a continuous-time framework, and we follow their approach. It is possible to incorporate hyperbolic discounting in a tractable way into the general model of Section 2.2. To illustrate its effects as clearly as possible, however, we instead incorporate it into the simpler model of Section 2.1.

Hyperbolic discounting is modeled by thinking of the agent as a sequence of different "selves," each of which exercises control at a different time. Specifically, from the vantage point of time 0, we divide the investor's horizon into two periods: a "present," which lasts until some random time $s > 0$; and a "future," which starts at time s. We think of the "present" as an interval during which control is exercised by the current self, and the "future" as an interval which is controlled by future selves. We assume that s follows a Poisson process with parameter ϕ, independent of other processes. At time s, a future self appears. That self's horizon can also be divided into a "present," which lasts until some random time, and a "future"; and so on.

Each self discounts utility which accrues within its "present" at $e^{-\delta t}$, and utility which accrues during its "future" at $\beta e^{-\delta t}$. The parameter $\beta < 1$ captures the idea that, under hyperbolic discounting, the present receives extra weight, relative to the future.

Hyperbolic discounting implies time-inconsistent behavior: the current self and future selves have different time preferences. The current self's beliefs about the actions of future selves are therefore important. Using the terminology of the literature, the current self can be "sophisticated," in that he correctly forecasts that future selves will use hyperbolic discounting; or he can be "naive," inaccurately believing that future selves will discount exponentially. For space reasons, we only analyze the naive case here. We expect the results for the sophisticated case to be qualitatively similar.

Consider a naive hyperbolic discounter who is holding stock at time t. If the next self arrives at some random time τ' in the future, the current self's decision problem is

$$\begin{aligned} N\left(W_t, B_t\right) \; = \; \max_{\tau \geq t} \;\; & E_t \left\{ e^{-\delta(\tau - t)} \left[u((1-k)W_\tau - B_\tau) I_{\{\tau < \tau'\}} + N\left((1-k)W_\tau, (1-k)W_\tau\right) I_{\{\tau < \tau'\}} \right. \right. \\ & \left. \left. + \widehat{N}(W_{\tau'}, B_{\tau'}) I_{\{\tau \geq \tau'\}} \right] \right\}. \end{aligned} \quad (35)$$

If the investor sells stock early enough, so that $\tau < \tau'$, he receives realization utility of $u((1-k)W_\tau - B_\tau)$ and the value function $N\left((1-k)W_\tau, (1-k)W_\tau\right)$. If, instead, $\tau > \tau'$, the current self receives the value function that he *thinks* will result from the actions of the future self. We use $\widehat{N}(W_t, B_t)$ to denote this perceived value function.

An important step is to note that

$$\widehat{N}(W_t, B_t) = \beta V(W_t, B_t). \quad (36)$$

Since the current self is a *naive* hyperbolic discounter, he thinks that future selves will use exponential discounting, and therefore that they will follow a strategy of selling once the gain reaches the liquidation point g_* derived in Section 2.1. The value function that the current self thinks will result from the actions of future selves is therefore $\beta V(W_t, B_t)$: the value function of an exponential discounter multiplied by β. The β factor appears because the current self discounts utility flows in the "future" period at $\beta e^{-\delta t}$ rather than at $e^{-\delta t}$.

In the Appendix, we prove:

Proposition 4: An investor with the decision problem in (35) will sell a position in stock once the gain $g_t = W_t/B_t$ reaches a liquidation point $g_{**} > 1$. The value function is $N(W_t, B_t) = B_t n(g_t)$, where

$$n(g_t) = \begin{cases} b_1 g_t^{\gamma_1} + b_2 g_t^{\gamma_2} & \text{if} \quad g_t < g_{**} \\ (1-k)g_t(1+n(1)) - 1 & \text{if} \quad g_t \geq g_{**} \end{cases}, \quad (37)$$

where γ_1 is given in (12),

$$\gamma_2 = \frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2\left(\delta + \phi\right)\sigma^2} - \left(\mu - \frac{1}{2}\sigma^2\right)\right] > 0 \quad (38)$$

$$b_1 = \frac{\phi \beta c_1}{\phi + \delta - \mu\gamma_1 - \frac{\sigma^2}{2}\gamma_1(\gamma_1 - 1)}, \quad (39)$$

where

$$c_1 = \frac{(1-k)}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)},$$

with g_* the unique solution to (11), and where b_2 and g_{**} are given by

$$\begin{aligned} b_1 g_{**}^{\gamma_1} + b_2 g_{**}^{\gamma_2} &= (1-k) g_{**}(1 + b_1 + b_2) - 1 \\ b_1 \gamma_1 g_{**}^{\gamma_1 - 1} + b_2 \gamma_2 g_{**}^{\gamma_2 - 1} &= (1-k)(1 + b_1 + b_2). \end{aligned}$$

Results

The top graphs in Figure 6 show how the liquidation point and initial utility $n(1)$ depend on the hyperbolic discounting parameter β. Here, we vary β while maintaining

$$(\mu, \sigma, \delta, k, \phi) = (0.03, 0.5, 0.08, 0.01, 3). \tag{40}$$

In particular, we set the arrival intensity of new selves to $\phi = 3$. The key finding is that, as we predicted, hyperbolic discounting makes the investor more impatient to realize gains: as β falls, g_{**} also falls.

The solid lines in the middle and bottom panels of Figure 6 show how the liquidation point g_{**} and initial utility $n(1)$ depend on the stock's expected return and standard deviation. When we vary μ or σ, we keep the remaining parameters fixed at their benchmark values, namely

$$(\mu, \sigma, \delta, k, \beta, \phi) = (0.03, 0.5, 0.08, 0.01, 0.9, 3). \tag{41}$$

Note that the benchmark value of β is 0.9.

The dotted lines in these graphs show what happens in the exponential discounting case, in other words, when $\beta = 1$ and $\phi = 0$. By comparing the solid and dotted lines, we see that hyperbolic discounting preserves the basic relationship between g_* on the one hand and μ and σ on the other; the effect is simply to shift the liquidation point down. It also preserves the basic relationship between $U(1)$ and μ and σ; it simply shifts the utility level down.

4 Applications

Our model may be helpful for thinking about a wide range of financial phenomena. We now discuss some of these potential applications. We divide the applications into those that relate to the trading behavior of investors (Section 4.1); and those that relate to asset pricing (Section 4.2). In Section 4.3, we discuss some of the model's testable predictions.

4.1 Investor trading behavior

The disposition effect

The disposition effect is the finding that individual investors have a greater propensity to sell stocks that have gone *up* in value since purchase, rather than stocks that have gone down (Odean, 1998). This fact has turned out to be something of a puzzle: the most obvious explanations fail to explain important features of the data. Consider, for example, the most obvious explanation of all, the "informed trading" hypothesis. Under this view, investors sell stocks that have gone up in value because they have private information that these stocks will subsequently fall, and they hold on to stocks that have gone down in value because they have private information that these stocks will subsequently rebound. The difficulty with this explanation, as Odean (1998) points out, is that the prior winners people sell subsequently do *better*, on average, than the prior losers they hold on to. Odean (1998) also considers other potential explanations based on taxes, rebalancing, and transaction costs, but argues that all of them fail to capture important aspects of the data.

Our analysis shows that a model that combines realization utility with a positive time discount factor predicts a strong disposition effect. In fact, unless forced to sell by a liquidity shock, the investor in our model *only* sells stocks trading at a gain, never a stock trading at a loss.

In simple two-period settings, Shefrin and Statman (1985) and Barberis and Xiong (2006) show that realization utility, with no time discounting but a prospect theory functional form for utility, can predict a disposition effect. This paper proposes a related, but distinct view of the disposition effect, namely that it arises from realization utility coupled with a *linear* functional form for utility and a *positive* time discount factor.

We emphasize that realization utility does not, on its own, predict a disposition effect. In other words, it is not enough to assume that the investor derives pleasure from realizing a gain and pain from realizing a loss. We need an additional ingredient in order to explain why the investor would want to realize a gain *today*, rather than hold out for the chance of realizing an even bigger gain tomorrow. Shefrin and Statman (1985) and Barberis and Xiong (2006) point out one possible extra ingredient: a prospect theory functional form for utility, and, in particular, a value function that is concave over gains and convex over losses. Such a functional form indeed explains the expediting of gains and the postponement of losses. Here, we propose an alternative extra ingredient: a sufficiently positive time discount factor.

Our model is also well-suited for thinking about the disposition-type effects that have been uncovered in other settings. Genesove and Mayer (2001), for example, find that homeowners are reluctant to sell their houses at prices below the original purchase price; and Heath, Huddart, and Lang (1999) find that executives are more likely to exercise stock options when the underlying stock price exceeds a reference point – the stock's highest price over the previous year – than when it falls below that reference point. Our analysis shows that a model that combines linear realization utility with a positive time discount rate can capture this evidence very easily.

Weber and Camerer (1995) provide some useful experimental evidence for the realization utility view of the disposition effect. In a laboratory setting, they ask subjects to trade six stocks over a number of periods. In each period, each stock can either go up or down. The six stocks have different probabilities of going up in any period, ranging from 0.35 to 0.65, but subjects are not told which stock is associated with each possible up-move probability.

Weber and Camerer (1995) find that, just as in field data, their subjects exhibit a disposition effect: they have a greater propensity to sell stocks trading at a gain relative to purchase price, rather than stocks trading at a loss. To try to understand the source of the effect, the authors consider an additional experimental condition in which the experimenter liquidates subjects' holdings, and then tells them that they are free to reinvest the proceeds in any way they like. If subjects were holding on to their losing stocks because they thought that these stocks would rebound, we would expect them to re-establish their positions in these losing stocks. In fact, subjects do *not* re-establish these positions. This casts doubt on belief-based explanations of the disposition effect, and lends support to the realization utility view, namely that subjects were refusing to sell their losers simply because it would have been painful to do so. Under this view, subjects were relieved when the experimenter intervened and did it for them.

Excessive trading

Using a large database of trading activity at a discount brokerage firm, Barber and Odean (2000) show that, before transaction costs, the average return of the individual investors in their sample is on par with a range of benchmarks; but that, after transaction costs, it falls *below* the benchmark returns. This last finding is puzzling: Why do people trade so much, when their trading activity hurts their performance? Barber and Odean (2000) consider a number of potential explanations, including taxes, rebalancing, and liquidity needs, but conclude that none of them can fully explain the patterns they observe.

Our model offers a simple explanation for this post-transaction-cost underperformance. From the perspective of investors, the underperformance is compensated by the occasional bursts of positive utility they experience when they realize gains.

It is straightforward to compute the probability that, over any interval after he first establishes a position in stock, the investor in our model trades at least once. This is not the same thing as a turnover rate, but it is related, and can therefore help us compare the trading frequency predicted by our model with that observed in actual brokerage accounts. When the investor first establishes a position, $g_0 = 1$. When g_t passes an upper barrier g_*, he liquidates the position. To compute the probability that the investor, after establishing a position, trades at least once in the s periods thereafter, we simply need to compute the probability that g_t passes g_* in the period $(0, s)$.

Proposition 5: The probability that at least one trade occurs in $(0, s)$ is:

$$\begin{aligned} G(s) &= N\left(\frac{-\ln g_* + \left(\mu - \frac{\sigma^2}{2}\right) s}{\sigma\sqrt{s}}\right) \\ &\quad + \exp\left(\frac{2\left(\mu - \frac{\sigma^2}{2}\right)\ln g_*}{\sigma^2}\right) N\left(\frac{-\ln g_* - \left(\mu - \frac{\sigma^2}{2}\right) s}{\sigma\sqrt{s}}\right). \end{aligned} \tag{42}$$

Proof of Proposition 5: See the Appendix.

Figure 7 shows how the probability of at least one trade in a stock over the year after it is bought, $G(1)$, depends on the model parameters. Some of the results are not surprising. As the investor becomes more impatient – as the exponential discount rate δ goes up, or as the hyperbolic discount rate β goes down – the probability of a trade goes up. And as transaction costs fall, the probability of a trade goes up.

The graphs with μ and σ on the horizontal axis are less predictable. In both cases, there are two factors at work. On the one hand, for any fixed liquidation point g_*, a higher μ or σ raises the likelihood that g_* will be reached. However, as we saw in several previous figures, the liquidation point g_* itself goes up as μ and σ go up, thereby lowering the chance that g_* will be reached. Without computing $G(1)$ explicitly, we cannot tell which factor will dominate.

Figure 7 shows that, interestingly, a different factor dominates in each of the two cases. As μ rises, the probability of a trade *falls*. Roughly speaking, as μ rises, the liquidation point rises more quickly than the stock's ability to catch it. As σ rises, however, the probability of a trade goes *up*: in this case, the liquidation point rises *less* quickly than the stock's ability to catch it.

The graphs show that, for our benchmark parameters, the model predicts a trading frequency that is not dissimilar to the turnover rates reported by Barber and Odean (2000) for brokerage accounts. When $\sigma = 50\%$, for example, the probability that an investor will trade a specific stock in his portfolio within a year of purchase is about 50%.

Underperformance *before* transaction costs

Some studies find that individual investors underperform benchmarks even before transaction costs (Odean, 1999). Our model may be able to shed light on this. The key insight is that, in our model, the investor *is* willing to buy a stock with a negative return premium, so long as the stock's volatility is sufficiently high. The reason is that, if the stock is volatile enough, it offers the chance of a sizeable gain, which the investor can enjoy realizing. Of course, a negative expected return stock can also fall in value. But the investor does not voluntarily realize losses, so this outcome only brings him disutility in the event of a liquidity shock. So long as the intensity of liquidity shocks is not too high, the investor is willing to invest in a negative expected return stock if its standard deviation is sufficiently high.

Note that, for the investor to be willing to buy a stock with a negative average return, it is not necessary that he literally say to himself, in advance of buying the stock: "Well, the stock could go down, but that's OK – I just won't sell it, so it won't hurt." Explicit reasoning of this kind may not be plausible. A more plausible way in which our prediction may manifest itself is through "remembered" utility. Suppose that, over time, the investor experiments with a variety of different stocks, including some with negative average returns. If, later on, he thinks back on his experience with these negative premium stocks, he will, on average, have good memories of them: in some cases, the stocks will have done well, and he will have sold them at a gain, thereby enjoying positive utility; in other cases, the stocks will have done poorly, but, since he did not voluntarily sell these stocks, they only led to disutility in the event of a liquidity shock. If liquidity shocks are not too frequent, his overall recalled experience of the negative premium stocks will be favorable, and so he may be happy to keep buying them.

Trading in up and down markets

Our model suggests a reason for the high overall level of trading activity, but also for why there is more trading in up markets than in down markets. First, our investor has a much greater propensity to *sell* stocks in a rising market: a rising market gives him more opportunities to realize gains, which is something he enjoys doing. This immediately implies that he will also have a much greater propensity to *buy* in a rising market. In order to buy a stock, our investor needs capital. To free up capital, he needs to sell his holdings of other stocks. But he will only be willing to do this in a rising market, not in a bear market.

4.2 Asset pricing

Our model has interesting implications not only for investor behavior but for asset pricing as well. We illustrate these implications using the simplest possible pricing model: one in which the economy is made up of homogeneous realization utility investors.

Specifically, consider an economy with a risk-free asset and N risky stocks, indexed by $i \epsilon \{1, \ldots N\}$. The risk-free asset is in perfectly elastic supply and earns a net return of zero. Stock i follows the price process

$$\frac{dS_i}{S_i} = \mu_i dt + \sigma_i dZ_{i,t},$$

where σ_i is constant over time. We assume, for now, that μ_i is also constant over time, and confirm this assumption later.

The economy contains a continuum of realization utility investors. At each time $t \geq 0$, each investor must either allocate all of his wealth to the risk-free asset, or all of his wealth to one of the stocks. We allow for transaction costs, search intervals, liquidity shocks, and loss aversion. Any investor holding a stock at time t therefore has the decision problem in (25), with $u(\cdot)$ given in (24). We assume that investors are homogeneous, so that δ, T, ρ, and λ are the same for all investors. Transaction costs, however, can differ across stocks.

In this economy, a sufficient condition for equilibrium is $U(1) = 0$, where $U(\cdot)$ is given in (26). If $U(1) = 0$, each investor is indifferent, at time 0, between holding the risk-free asset and allocating his wealth to a stock. By assigning some investors to each stock and the rest to the risk-free asset, we can therefore clear markets at time 0. If, at any point in the future, some investors sell their holdings of stock i – whether because of a liquidity shock or because, for these investors, the stock has reached its liquidation point – we can reassign some investors from the risk-free asset to the stock, thereby again clearing markets. Since the investor-specific parameters δ, T, ρ, and λ are constant, for any particular stock, the equation $U(1) = 0$ is always satisfied by the same μ_i, so that μ_i is indeed constant over time, as we assumed earlier.

With this structure in place, we now discuss two of our model's more interesting asset pricing implications.

The negative volatility premium

Ang et al. (2005) show that, in the cross-section, and after controlling for well-known predictors of cross-sectional returns, a stock's daily return volatility over the previous year *negatively* predicts its return in the following year: highly volatile stocks subsequently earn lower average returns.

This is a puzzling finding. Even if we allow ourselves to think of a stock's own volatility as risk, the result is the opposite of what we would expect: it says that "riskier" stocks have *lower* average returns. Nor can the result be fully explained using a model that combines differences of opinion with short-sale constraints: the effect persists even after controlling for differences of opinion using dispersion in analyst forecasts.

Our model offers a novel explanation. The key insight comes from the middle-right graph in Figure 2: the finding that, holding other parameters constant, the greater a stock's volatility, the higher the investor's initial utility. This result suggests that, in an equilibrium context, highly volatile stocks may experience heavy buying pressure from investors who care about realization utility. These stocks may therefore become overpriced, and, as a result, may earn low average returns.

We can check this in our simple equilibrium model. We assign all investors the same benchmark parameters

$$(\delta, T, \rho, \lambda) = (0.08, 0.1, 0.1, 1.5), \tag{43}$$

and suppose that the transaction cost parameter is the same for all stocks, namely $k = 0.01$. For a range of values of σ, we solve $U(1) = 0$ to obtain the equilibrium expected return that a stock with any particular standard deviation must earn.

The top-left graph in Figure 8 plots the resulting relationship between expected return and standard deviation. The graph confirms our prediction: more volatile stocks earn lower average returns; in this sense, they are overpriced.

A counterfactual prediction of the top-left graph is that the average equity premium is negative. One way to obtain a negative relationship between expected return and volatility in conjunction with a *positive* equity premium is to suppose that investors apply different decision rules to different components of their wealth. In particular, suppose that investor use a standard concave utility function to allocate most of their wealth between a risk-free asset and a stock market index; but that, for the remainder of their wealth – the "play" money in their brokerage accounts which they allocate across individual stocks – realization utility preferences apply. The combination of a concave utility function on the one hand and realization utility on the other may be able to reconcile a high equity premium with a negative relationship between expected return and standard deviation.

Heavy trading of overvalued assets

A robust empirical finding is that assets that are highly valued, and possibly overvalued, are also heavily traded (Hong and Stein, 2007). Growth stocks, for example, are more heavily traded than value stocks; the highly-priced internet stocks of the late 1990s changed hands at a rapid pace; and shares at the center of famous bubble episodes, such as those of the East India Company at the time of the South Sea bubble, also experienced heavy trading.

Our model may be able to explain this coincidence of high prices and heavy trading. Moreover, it predicts that this phenomenon should occur at times when the value of the underlying asset is especially uncertain.

Suppose that the uncertainty about an asset's value goes up, pushing up its standard deviation σ. As noted earlier, investors who care about realization utility will now find the asset more attractive. If there are many such investors in the economy, the asset's price may be pushed up.

At the same time, the top-right graph in Figure 7 shows that, as σ goes up, the probability that the investor will trade the asset *also* goes up: simply put, a more volatile stock will reach its liquidation point more rapidly. In this sense, the overvaluation will coincide with higher turnover, and this will occur when uncertainty about the underlying asset value is especially high. Under this view, the late 1990s were years where realization utility investors, attracted by the high uncertainty of technology stocks, bought these stocks, pushing their prices up; as (some of) these stocks rapidly reached their liquidation points, the realization utility investors sold them, and then immediately bought new ones.

We can illustrate this result using our simple equilibrium framework. As in our discussion of the negative volatility premium, we assign all individual investors the benchmark parameters in (43) and assume that the transaction cost parameter is the same for all stocks, namely $k = 0.01$. For a range of values of σ, we compute, as before, the equilibrium expected return the stock must earn, but also, as a guide to the intensity of trading, the probability of trade given in (42).

The top-right graph in Figure 8 plots the resulting relationship between expected return and trade probability. It confirms that stocks with lower expected returns – stocks that are "overpriced" – will experience more turnover.

4.3 Testable predictions

Our model addresses some puzzling facts, but also makes a number of new predictions. The most natural predictions emerge from Figure 7, which shows how the probability of trade depends on various parameters.

One of these predictions is not especially surprising: the investor trades more frequently when transaction costs are lower. Three other predictions, however, are more novel: The investor holds stocks with a higher average return for longer, before selling them. Stocks with higher volatility, however, are sold more quickly. And the more impatient the investor is, the more often he trades.

The prediction relating how long a stock is held to its average return is difficult to test because the average return perceived by individual investors may differ from the actual average return. Growth stocks, for example, have low average returns, but it is likely that some individual investors *perceive* them to have high average returns.

The prediction relating how long a stock is held to its volatility is easier to test. Indeed, after making this prediction, we found that the answer is already available in the literature. Zuckerman (2006) reports that the individual investors in the Barber and Odean (2000) database *do* hold more volatile stocks for shorter periods of time before selling them.

Our prediction relating trading frequency to investor impatience is harder to test, but by no means impossible. The difficulty here is obtaining an estimate of investor impatience. In recent years, researchers have pioneered clever techniques for extracting information about investors' psychological profiles. Grinblatt and Keloharju (2006), for example, use military test scores from Finland to estimate overconfidence. This success raises the possibility that a test of the link between impatience and trading frequency can also be implemented.

5 Conclusion

We study the possibility that, aside from standard sources of utility, investors also derive utility from *realizing* gains and losses on individual investments that they own. We propose a tractable model of this "realization utility," derive its predictions, and show that it can shed light on a number of puzzling facts. These include the poor trading performance of individual investors, the disposition effect, the greater turnover in up markets, the negative premium to volatility in the cross-section, and the heavy trading of highly valued assets. Underlying some of these applications is one of our model's more novel predictions: that, even if the form of realization utility is linear or concave, investors can be risk-*seeking*.

6 Appendix

Proof of Proposition 1: At time t, the investor can either liquidate his position immediately, or he can hold his position for an infinitesimal period dt, so that

$$V\left(W_t, B_t\right) = \max\left\{u\left((1-k)W_t - B_t\right) + V\left((1-k)W_t, (1-k)W_t\right), E_t\left[e^{-\delta dt} V\left(W_{t+dt}, B_{t+dt}\right)\right]\right\}. \tag{44}$$

We conjecture that the value function takes the form

$$V\left(W_t, B_t\right) = B_t U\left(g_t\right),$$

where $g_t = W_t/B_t$. Substituting this into (44) and noting that $B_{t+dt} = B_t$, we obtain

$$B_t U(g_t) = \max\left\{B_t u((1-k)g_t - 1) + (1-k)B_t g_t U(1), B_t E_t[e^{-\delta dt} U(g_{t+dt})]\right\}.$$

Cancelling the B_t factor from both sides gives

$$U\left(g_t\right) = \max\left\{u\left((1-k)g_t - 1\right) + (1-k)g_t U\left(1\right), \quad E_t\left[e^{-\delta dt} U\left(g_{t+dt}\right)\right]\right\}.$$

When the investor initially acquires a position in stock, $g_t = 1$. Until the stock is sold, g_t fluctuates according to

$$\frac{dg_t}{g_t} = \mu dt + \sigma dZ_t.$$

Ito's lemma then implies

$$E_t\left[e^{-\delta dt} U\left(g_{t+dt}\right)\right] = U\left(g_t\right) + \left[\frac{1}{2}\sigma^2 g_t^2 U''\left(g_t\right) + \mu g_t U'\left(g_t\right) - \delta U\left(g_t\right)\right] dt,$$

so that, using $u(x) = x$,

$$U\left(g_t\right) = \max\left\{(1-k)g_t\left(1 + U\left(1\right)\right) - 1, U\left(g_t\right) + \left[\frac{1}{2}\sigma^2 g_t^2 U''\left(g_t\right) + \mu g_t U'\left(g_t\right) - \delta U\left(g_t\right)\right] dt\right\}. \tag{45}$$

Equation (45) implies that any solution to (6) must satisfy

$$U\left(g_t\right) \geq (1-k)g_t\left(1 + U\left(1\right)\right) - 1 \tag{46}$$

and

$$\frac{1}{2}\sigma^2 g_t^2 U''\left(g_t\right) + \mu g_t U'\left(g_t\right) - \delta U\left(g_t\right) \leq 0. \tag{47}$$

Formally speaking, the decision problem in (6) is an optimal stopping problem. To solve it, we first construct a function $U(g_t)$ that satisfies conditions (46) and (47), and that is also continuously differentiable – this last condition is also known as the "smooth pasting" condition. We then verify that $U(g_t)$ does indeed solve problem (6).

We construct $U(g_t)$ in the following way. If g_t is low – specifically, if $g_t \epsilon (0, g_*)$ – we suppose that the investor continues holding his current position. In this "continuation" region, then, equation (45) is maximized by the second term within parentheses, so that (47) holds with equality. If g_t is sufficiently high – specifically, if $g_t \epsilon (g_*, \infty)$ – we suppose that the investor liquidates his position. In this "liquidation" region, equation (45) is maximized by the first term within parentheses, so that (46) holds with equality. We call g_* the "liquidation point" – the percentage gain in value, relative to the cost basis, at which the investor liquidates his holdings of a stock.

In the **continuation** region, $g_t \in (0, g_*)$, we therefore have

$$\frac{1}{2}\sigma^2 g_t^2 U''(g_t) + \mu g_t U'(g_t) - \delta U(g_t) = 0.$$

The solution to this equation is

$$U(g_t) = c_1 g_t^{\gamma_1} + c_2 g_t^{\gamma_2},$$

where γ_1 and γ_2 are the roots of

$$\frac{1}{2}\sigma^2\gamma^2 + \left(\mu - \frac{1}{2}\sigma^2\right)\gamma - \delta = 0,$$

namely

$$\gamma_1 = \frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2\delta\sigma^2} - \left(\mu - \frac{1}{2}\sigma^2\right)\right] > 0 \tag{48}$$

$$\gamma_2 = -\frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2\delta\sigma^2} + \left(\mu - \frac{1}{2}\sigma^2\right)\right] < 0. \tag{49}$$

The assumption that $\delta > \mu$ implies that $\gamma_1 > 1$. Since the value function must be bounded as g_t goes to zero, c_2 must equal zero. The value function in the continuation region is therefore

$$U(g_t) = c_1 g_t^{\gamma_1}, \tag{50}$$

where c_1 is determined below.

In the **liquidation** region, $g_t \in (g_*, \infty)$, we have

$$U(g_t) = (1-k)g_t(1 + U(1)) - 1. \tag{51}$$

Note that the liquidation point g_* satisfies $g_* \geq 1$. For if $g_* < 1$, then $g_t = 1$ would fall into the liquidation region, which, from (51), would imply

$$U(1) = (1-k)U(1) - k.$$

For $k > 0$ and $U(1) > 0$, this is a contradiction. Since $g_* \geq 1$, then, we infer from (50) that $U(1) = c_1$.

The value function must be continuous and smooth around the liquidation point g_*. Thus,

$$\begin{aligned} c_1 g_*^{\gamma_1} &= (1-k) g_* (1+c_1) - 1 && (52) \\ \gamma_1 c_1 g_*^{\gamma_1 - 1} &= (1-k)(1+c_1). && (53) \end{aligned}$$

Substituting out c_1, we obtain the following equation for g_*:

$$f(g_*) = \frac{1-k}{\gamma_1 - 1} g_*^{-(\gamma_1 - 1)} + (1-k) g_* - \frac{\gamma_1}{\gamma_1 - 1} = 0. \quad (54)$$

It is straightforward to check that $f(1) = \frac{-k\gamma_1}{\gamma_1 - 1} < 0$, that $f(\infty) > 0$, and that $f(\cdot)$ is a convex function in $(1, \infty)$. Equation (54) therefore has a unique solution $g_* > 1$.

From equation (52),

$$c_1 = \frac{1-k}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)}. \quad (55)$$

The investor's value function in the continuation region is therefore

$$U(g_t) = c_1 g_t^{\gamma_1} = \frac{(1-k) g_t^{\gamma_1}}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)}.$$

We now verify that the constructed value function is indeed optimal. Substituting $V(W_t, B_t) = B_t U(g_t)$ into (6) and cancelling the B_t factor reduces the stopping problem to

$$U(g_t) = \max_{\tau \geq t} E_t \left\{ e^{-\delta(\tau - t)} \left[u\left((1-k) g_\tau - 1\right) + (1-k) g_\tau U(1) \right] \right\}.$$

We first verify that the function $U(g_t)$ in the statement of the theorem satisfies conditions (46) and (47). Define

$$f_1(g_t) \equiv \frac{(1-k) g_t^{\gamma_1}}{\gamma_1 g_*^{\gamma_1 - 1} - (1-k)}$$

and

$$f_2(g_t) \equiv (1-k) g_t (1 + U(1)) - 1.$$

By construction $f_1(g_*) = f_2(g_*)$. Since $\gamma_1 > 1$, $f_1(g_t)$ is convex, so that $f_1(g_t) > f_2(g_t)$ for $g_t \in (0, g_*)$. Thus, condition (46) always holds. Furthermore, by construction,

$$\frac{1}{2} \sigma^2 g_t^2 U''(g_t) + \mu g_t U'(g_t) - \delta U(g_t) = 0$$

for $g_t < g_*$. For $g_t \geq g_*$,

$$\frac{1}{2} \sigma^2 g_t^2 U''(g_t) + \mu g_t U'(g_t) - \delta U(g_t) = (\mu - \delta) U(g_t) < 0.$$

Thus, condition (47) also holds.

Note also that $U(g_t)$ has an increasing derivative in $(0, g_*)$ and a derivative of $(1-k)(1+U(1))$ in (g_*, ∞). U' is therefore bounded.

For any stopping time τ, Ito's lemma for twice-differentiable functions with absolutely continuous first derivatives – see, for example, Revuz and Yor (1999), Chapter 6 – implies

$$e^{-\delta(\tau-t)}U(g_\tau) = U(g_t) + \int_t^\tau \left[\frac{1}{2}\sigma^2 g_s^2 U''(g_s) + \mu g_s U'(g_s) - \delta U(g_s)\right] ds + \int_t^\tau \sigma g_s U'(g_s)\, ds.$$

From condition (47), the first integral is non-positive, while the bound on U' implies that the second integral is a martingale. We therefore have

$$E_t\left[e^{-\delta(\tau-t)}U(g_\tau)\right] \le U(g_t).$$

Furthermore, condition (46) implies

$$E_t\left\{e^{-\delta(\tau-t)}\left[((1-k)g_t - 1) + (1-k)g_t U(1)\right]\right\} \le E_t\left[e^{-\delta(\tau-t)}U(g_\tau)\right].$$

Combining these two inequalities, we obtain that, for any stopping time τ,

$$E_t\left\{e^{-\delta(\tau-t)}\left[((1-k)g_t - 1) + (1-k)g_t U(1)\right]\right\} \le U(g_t).$$

The constructed value function $U(g_t)$ is therefore at least as good as the value function generated by any alternative strategy.

Proof of Proposition 2: Proposition 2 is the special case of Proposition 3 in which $\lambda = 1$.

Proof of Proposition 3: At time t, the investor can either liquidate his position, or he can hold it for an infinitesimal period dt, so that:

$$\begin{aligned} V(W_t, B_t) &= \max \begin{array}{c} \{u((1-k)W_t - B_t) + e^{-\rho T}e^{-\delta T}V((1-k)W_t, (1-k)W_t), \\ (1-\rho dt)E_t[e^{-\delta dt}V(W_{t+dt}, B_{t+dt})] + \rho dt\,[u((1-k)W_t - B_t)]\} \end{array} \quad (56) \\ &= \max \begin{array}{c} \{u((1-k)W_t - B_t) + e^{-(\rho+\delta)T}V((1-k)W_t, (1-k)W_t), \\ E_t\left[e^{-\delta dt}V(W_{t+dt}, B_{t+dt})\right] + \rho dt\,[u((1-k)W_t - B_t) - V(W_t, B_t)]\} \end{array} \quad (57) \end{aligned}$$

The first expression within the parentheses in (56) shows what happens if the investor liquidates a stock today. First, he receives utility $u((1-k)W_t - B_t)$ from the realized gain or loss. He then needs to hold the proceeds in cash for T periods. With probability $e^{-\rho T}$, there is no liquidity shock during the search interval, and his future value function is simply $e^{-\delta T}V((1-k)W_t, (1-k)W_t)$: the value function at the end of the search interval, discounted back. With probability $1 - e^{-\rho T}$, there *is* a liquidity shock during the search interval, and the investor is forced to take his money out of the brokerage account. Since the investor is already in cash, this entails utility of zero. Combining the two outcomes gives the first expression within the parentheses.

The second expression within the parentheses shows what happens if the investor decides to keep holding his position for an infinitesimal period dt. With probability $e^{-\rho dt} \approx 1 - \rho dt$, there is no liquidity shock, and the investor's value function is simply the expected future value function, discounted back. With probability $1 - e^{-\rho dt} \approx \rho dt$, there *is* a liquidity shock, and the investor sells his holdings and exits, which entails utility of $u((1-k)W_t - B_t)$.

We conjecture that the value function takes the form

$$V\left(W_t, B_t\right) = B_t U\left(g_t\right).$$

Substituting this into (57), cancelling the B_t factor from both sides, and applying Ito's lemma gives

$$U\left(g_t\right) = \max \begin{array}{c} \{u((1-k)g_t - 1) + e^{-(\rho+\delta)T}(1-k)g_t U\left(1\right), \\ \left(1-\delta dt\right) U\left(g_t\right) + \left[\mu g_t U'\left(g_t\right) + \frac{1}{2}\sigma^2 g_t^2 U''\left(g_t\right)\right] dt + \rho dt\left[u((1-k)g_t - 1) - U\left(g_t\right)\right]\} \end{array}. \tag{58}$$

As before, we conjecture that there are two regions: a continuation region, $g_t \epsilon (0, g_*)$, and a liquidation region, $g_t \epsilon (g_*, \infty)$. In the **continuation** region,

$$\frac{1}{2}\sigma^2 g_t^2 U''\left(g_t\right) + \mu g_t U'\left(g_t\right) - \left(\rho + \delta\right) U\left(g_t\right) + \rho u\left((1-k)g_t - 1\right) = 0. \tag{59}$$

The form of the $u(\cdot)$ term depends on whether its argument, $(1-k)g_t - 1$, is greater or less than zero. Note that the cross-over point, $g_t = \frac{1}{1-k}$, satisfies $g_* \geq \frac{1}{1-k}$. For if $g_* < \frac{1}{1-k}$, then $g_t = \frac{1}{1-k}$ would be in the liquidation region, which, from (58), would imply

$$U\left(\frac{1}{1-k}\right) = e^{-(\rho+\delta)T} U(1).$$

If either $\rho > 0$ or $\delta > 0$, this contradicts the plausible restriction that $U(g_t)$ be increasing in g_t. Since $g_* \geq \frac{1}{1-k}$, we further subdivide the continuation region $(0, g_*)$ into two subregions, $(0, \frac{1}{1-k})$, and $(\frac{1}{1-k}, g_*)$.

For $g_t \in (0, \frac{1}{1-k})$, equation (59) becomes

$$\frac{1}{2}\sigma^2 g_t^2 U''\left(g_t\right) + \mu g_t U'\left(g_t\right) - \left(\rho + \delta\right) U\left(g_t\right) + \rho\lambda\left((1-k)g_t - 1\right) = 0.$$

The solution to this equation is

$$U\left(g_t\right) = c_1 g_t^{\gamma_1} + \frac{\rho\lambda(1-k)}{\rho + \delta - \mu} g_t - \frac{\rho\lambda}{\rho + \delta} \quad \text{for} \quad g_t \in (0, \frac{1}{1-k}), \tag{60}$$

where γ_1 is defined in equation (17), and where c_1 is determined below.

For $g_t \in (\frac{1}{1-k}, g_*)$, equation (59) becomes

$$\frac{1}{2}\sigma^2 g_t^2 U''(g_t) + \mu g_t U'(g_t) - (\rho+\delta) U(g_t) + \rho((1-k)g_t - 1) = 0.$$

The solution to this equation is

$$U(g_t) = b_1 g_t^{\gamma_1} + b_2 g_t^{\gamma_2} + \frac{\rho(1-k)}{\rho+\delta-\mu} g_t - \frac{\rho}{\rho+\delta} \quad \text{for} \quad g_t \in \left(\frac{1}{1-k}, g_*\right),$$

where

$$\gamma_2 = -\frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2(\rho+\delta)\sigma^2} + \left(\mu - \frac{1}{2}\sigma^2\right)\right] < 0,$$

and where b_1 and b_2 are determined below.

The value function must be continuous and smooth around $g_t = \frac{1}{1-k}$. We therefore have

$$\begin{aligned} c_1\left(\frac{1}{1-k}\right)^{\gamma_1} &= b_1\left(\frac{1}{1-k}\right)^{\gamma_1} + b_2\left(\frac{1}{1-k}\right)^{\gamma_2} - \frac{(\lambda-1)\mu\rho}{(\rho+\delta-\mu)(\rho+\delta)} \\ c_1\gamma_1\left(\frac{1}{1-k}\right)^{\gamma_1-1} &= b_1\gamma_1\left(\frac{1}{1-k}\right)^{\gamma_1-1} + b_2\gamma_2\left(\frac{1}{1-k}\right)^{\gamma_2-1} - \frac{(\lambda-1)(1-k)\rho}{\rho+\delta-\mu}. \end{aligned}$$

These equations imply

$$b_2 = \frac{1}{(\gamma_1-\gamma_2)} \frac{(\lambda-1)\rho(1-k)^{\gamma_2}(\mu\gamma_1-\rho-\delta)}{(\rho+\delta-\mu)(\rho+\delta)}.$$

In the **liquidation** region, using the fact that $g_* \geq 1$, we have

$$U(g_t) = (1-k)g_t(1 + e^{-(\rho+\delta)T} U(1)) - 1.$$

The value function must also be continuous and smooth around the liquidation point g_*. Thus,

$$\begin{aligned} b_1 g_*^{\gamma_1} + b_2 g_*^{\gamma_2} + \frac{\rho(1-k)}{\rho+\delta-\mu} g_* &= (1-k)g_*(1 + e^{-(\rho+\delta)T} U(1)) - \frac{\delta}{\rho+\delta} \\ b_1\gamma_1 g_*^{\gamma_1-1} + b_2\gamma_2 g_*^{\gamma_2-1} + \frac{\rho(1-k)}{\rho+\delta-\mu} &= (1-k)(1 + e^{-(\rho+\delta)T} U(1)). \end{aligned}$$

Since, from equation (60),

$$U(1) = c_1 + \frac{\rho\lambda(\mu - k\rho - k\delta)}{(\rho+\delta)(\rho+\delta-\mu)},$$

we further obtain

$$b_1 g_*^{\gamma_1} + b_2 g_*^{\gamma_2} + (1-k) g_* \frac{\mu - \delta}{\rho + \delta - \mu} + \frac{\delta}{\delta + \rho} = (1-k) g_* e^{-(\rho+\delta)T} \left(c_1 + \frac{\rho\lambda(\mu - k\rho - k\delta)}{(\rho+\delta)(\rho+\delta-\mu)} \right)$$

and

$$(\gamma_1 - 1)\, b_1 g_*^{\gamma_1} + (\gamma_2 - 1)\, b_2 g_*^{\gamma_2} = \frac{\delta}{\rho + \delta}.$$

Proof of Proposition 4: At time t, the investor can either liquidate his position, or he can hold it for an infinitesimal period dt, so that:

$$\begin{aligned} N(W_t, B_t) &= \max\{(1-k)W_t - B_t + N((1-k)W_t, (1-k)W_t), \\ &\quad e^{-\phi dt} E_t[e^{-\delta dt} N(W_{t+dt}, B_{t+dt})] + (1 - e^{-\phi dt}) E_t[e^{-\delta dt} \widehat{N}(W_{t+dt}, B_{t+dt})]\}. \end{aligned} \tag{61}$$

If the current self sells stock now, he receives realized utility of $(1-k)W_t - B_t$ and a cash balance of $(1-k)W_t$. Alternatively, he may continue to hold his stock position for an infinitesimal period dt. With probability $e^{-\phi dt}$, the new self does *not* arrive during this interval, in which case the current self receives the discounted expected value function $E_t[e^{-\delta dt} N(W_{t+dt}, B_{t+dt})]$. With probability $1 - e^{-\phi dt}$, the new self *does* arrive during this interval, in which case the current self receives the discounted expected value function that he *thinks* will result from the actions of the future self.

We conjecture that

$$N(W_t, B_t) = B_t n(g_t).$$

Substituting this and $\widehat{N}(W_t, B_t) = \beta V(W_t, B_t)$ into (61), and cancelling the B_t factor leads to

$$n(g_t) = \max\left\{(1-k)g_t - 1 + (1-k)g_t n(1), e^{-(\phi+\delta)dt} E_t(n(g_{t+dt})) + (1 - e^{-\phi dt}) e^{-\delta dt} \beta E_t(U(g_t))\right\}.$$

Applying Ito's lemma, we obtain

$$\begin{aligned} n(g_t) &= \max\{(1-k)g_t - 1 + (1-k)g_t n(1), \\ &\quad n(g_t) + dt\left[\frac{1}{2}\sigma^2 g_t^2 n''(g_t) + \mu g_t n'(g_t) - (\phi+\delta) n(g_t) + \phi\beta U(g_t)\right]\}. \end{aligned}$$

As before, we conjecture that there are two regions: a continuation region, $g_t \,\epsilon\, (0, g_{**})$, where the current self keeps holding the stock, and a liquidation region, $g_t \,\epsilon\, (g_{**}, \infty)$, where the current self sells his position.

In the **continuation** region,

$$\frac{1}{2}\sigma^2 g_t^2 n''(g_t) + \mu g_t n'(g_t) - (\phi+\delta) n(g_t) + \phi\beta U(g_t) = 0. \tag{62}$$

We conjecture that $g_{**} < g_*$, where g_* is the liquidation point in the *exponential* discounting model of Section 2.1. From (50), this means that

$$U(g_t) = c_1 g_t^{\gamma_1}, \qquad g_t \; \epsilon \; (0, g_{**}),$$

where γ_1 and c_1 are given in (48) and (55), respectively. The solution to (62) is then

$$n(g_t) = b_1 g_t^{\gamma_1} + b_2 g_t^{\gamma_2}, \tag{63}$$

where

$$\begin{aligned} b_1 &= \frac{\phi \beta c_1}{\phi + \delta - \mu \gamma_1 - \frac{\sigma^2}{2}\gamma_1(\gamma_1 - 1)} \\ \gamma_2 &= \frac{1}{\sigma^2}\left[\sqrt{\left(\mu - \frac{1}{2}\sigma^2\right)^2 + 2(\delta + \phi)\sigma^2} - \left(\mu - \frac{1}{2}\sigma^2\right)\right] > 0, \end{aligned}$$

and where b_2 is determined below.

In the **liquidation** region, $g_t \; \epsilon \; (g_{**}, \infty)$, we have

$$n(g_t) = (1-k) g_t (1 + n(1)) - 1.$$

By the usual argument, $g_{**} \geq 1$, so that, from (63), $n(1) = b_1 + b_2$. At the liquidation point g_{**}, $n(\cdot)$ must be continuous and smooth. This implies

$$\begin{aligned} b_1 g_{**}^{\gamma_1} + b_2 g_{**}^{\gamma_2} &= (1-k) g_{**} (1 + b_1 + b_2) - 1 \\ b_1 \gamma_1 g_{**}^{\gamma_1 - 1} + b_2 \gamma_2 g_{**}^{\gamma_2 - 1} &= (1-k)(1 + b_1 + b_2). \end{aligned}$$

Proof of Proposition 5: Let τ denote the first time at which the process g_t passes the upper barrier g_*. Since g_t follows a geometric Brownian motion, its logarithm follows a Brownian motion:

$$d \ln(g_t) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dZ_t.$$

Then, τ is also the first time at which the process $\ln(g_t)$ passes the level $\ln(g_*)$. Let $G(\tau; g_0)$ denote the distribution of τ conditional on g_0. Following Ingersoll (1988, pp. 353), but in our notation,

$$\begin{aligned} G(\tau; g_0) &= N\left(\frac{\ln(g_0/g_*) + (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right) \\ &\quad + \exp\left[-\frac{2(\mu - \sigma^2/2)\ln(g_0/g_*)}{\sigma^2}\right] N\left(\frac{\ln(g_0/g_*) - (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right). \end{aligned}$$

7 References

Barber, B. and T. Odean (2000), "Trading is Hazardous to Your Wealth: The Common Stock Performance of Individual Investors," *Journal of Finance* 55, 773-806.

Barberis, N. and W. Xiong (2006), "What Drives the Disposition Effect? An Analysis of a Long-standing Preference-based Explanation," Working paper, Yale University.

Frazzini, A. (2006), "The Disposition Effect and Underreaction to News," *Journal of Finance* 61, 2017-2046.

Genesove, D. and C. Mayer (2001), "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics* 116, 1233-1260.

Grenadier, S. and N. Wang (2007), "Investment under Uncertainty and Time-Inconsistent Preferences," *Journal of Financial Economics* 84, 2-39.

Grinblatt, M. and B. Han (2005), "Prospect Theory, Mental Accounting, and the Disposition Effect," *Journal of Financial Economics* 78, 311-339.

Grinblatt, M. and M. Keloharju (2006), "Sensation-seeking, Overconfidence, and Trading Activity," Working paper, UC Los Angeles.

Gross, L. (1982), *The Art of Selling Intangibles: How to Make Your Million($) by Investing Other People's Money*, New York Institute of Finance: New York.

Harris, C. and D. Laibson (2004), "Instantaneous Gratification," Working paper, Harvard University.

Heath, C., Huddart, S., and M. Lang (1999), "Psychological Factors and Stock Option Exercise," *Quarterly Journal of Economics* 114, 601-627.

Hong, H. and J. Stein (2007), "Disagreement and the Stock Market," *Journal of Economic Perspectives* 12, 109-128.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263-291.

Kyle, A., Ou-Yang, H., and W. Xiong (2006), "Prospect Theory and Liquidation Decisions," *Journal of Economic Theory* 129, 273-288.

Odean, T. (1998), "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53, 1775-1798.

Odean, T. (1999), “Do Investors Trade Too Much?” *American Economic Review* 89, 1279-1298.

Revuz, D. and M. Yor (1999), *Continuous Martingales and Brownian Motion*, Springer: New York.

Shefrin, H. and M. Statman (1985), “The Disposition to Sell Winners too Early and Ride Losers too Long,” *Journal of Finance* 40, 777-790.

Tversky, A. and D. Kahneman (1992), “Advances in Prospect Theory: Cumulative Representation of Uncertainty,” *Journal of Risk and Uncertainty* 5, 297-323.

Weber, M. and C. Camerer (1998), “The Disposition Effect in Securities Trading: An Experimental Analysis,” *Journal of Economic Behavior and Organization* 33, 167-184.

Zuckerman, R. (2006), “It’s not the Trader, It’s the Trade: A Cross-sectional Analysis of Individual Traders’ Activity,” Working paper, Yale University.

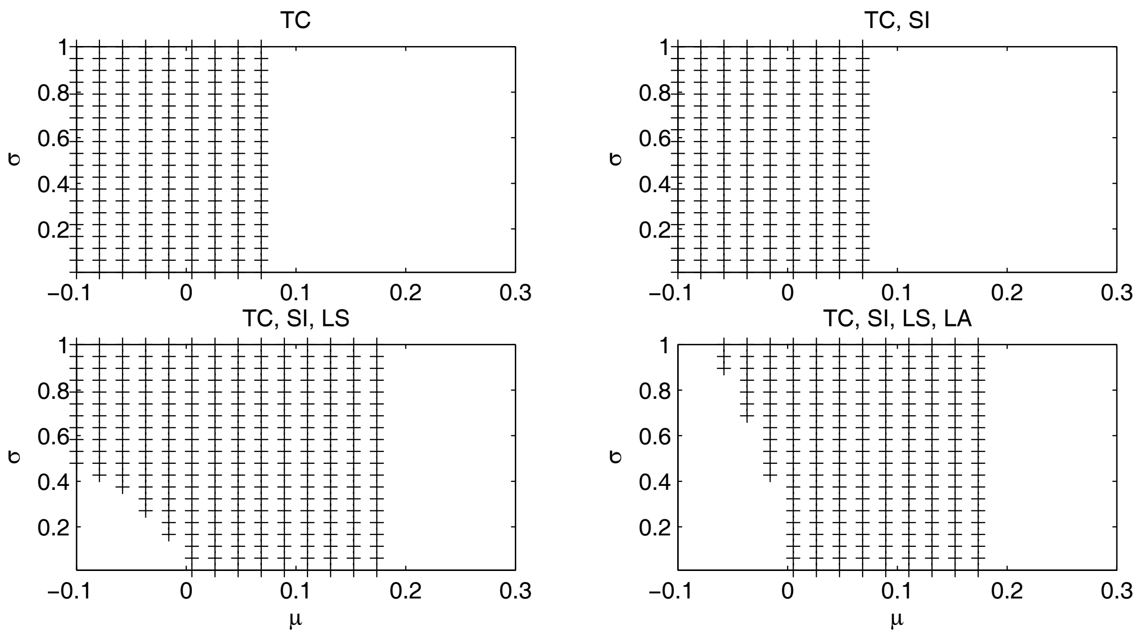

Figure 1. The graphs show, for an investor who derives utility from realized gains and losses, the range of values of a stock's expected return μ and standard deviation σ for which the investor is willing to buy the stock. The top-left graph corresponds to a model which allows for a transaction cost (TC); the top-right graph to a model which also allows for a search interval between the sale of one stock and the purchase of another stock (SI); the middle-left graph to a model which further allows for an exogeneous liquidity shock (LS); and the middle-right graph to a model which also allows for loss aversion (LA).

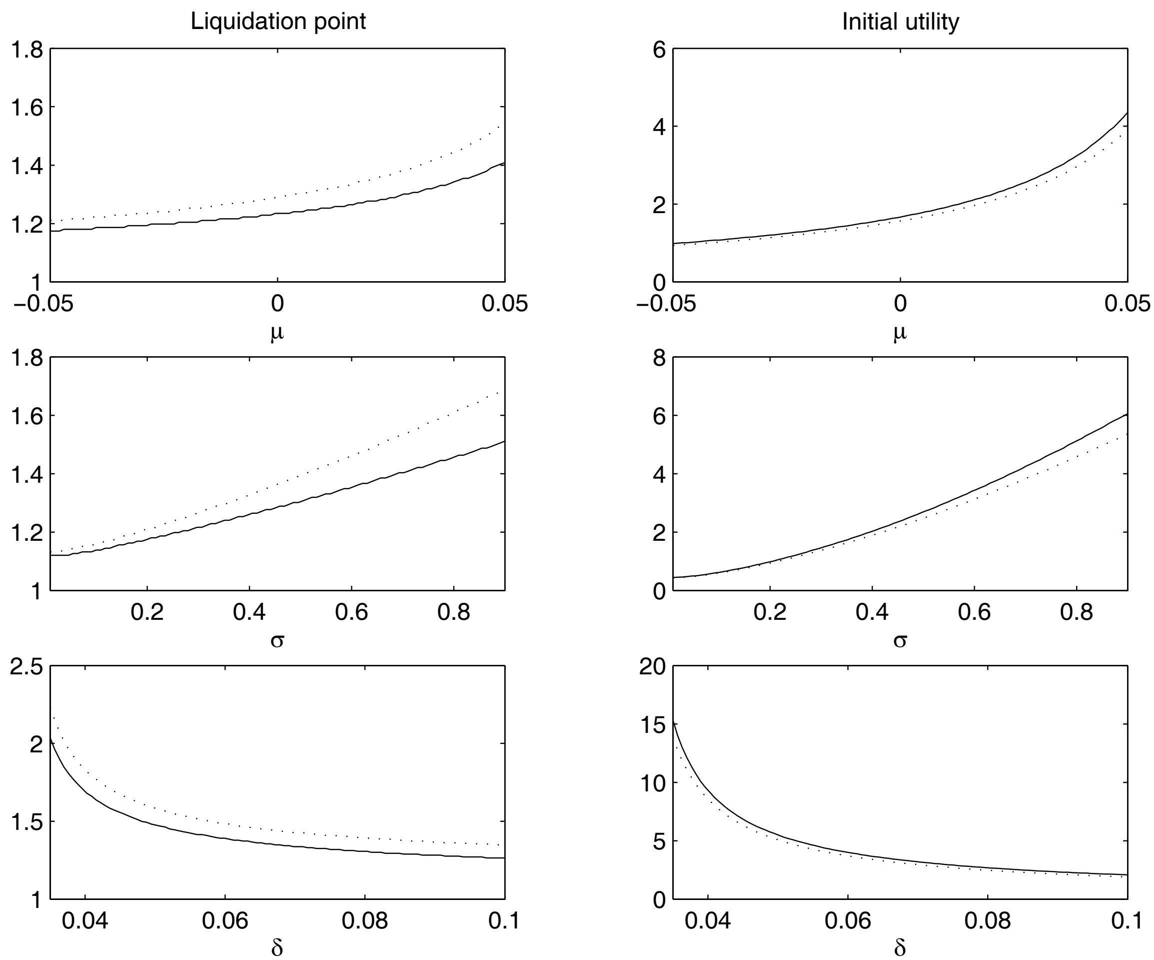

Figure 2. The graphs show, for an investor who cares about utility from realized gains and losses, how the liquidation point at which he sells a stock and the initial utility from buying it depend on the stock's expected return μ, its standard deviation σ, and the investor's discount rate δ. The solid lines correspond to a model that allows for a transaction cost. The dotted lines correspond to a model that allows for both a transaction cost and a search interval between the sale of one stock and the purchase of another.

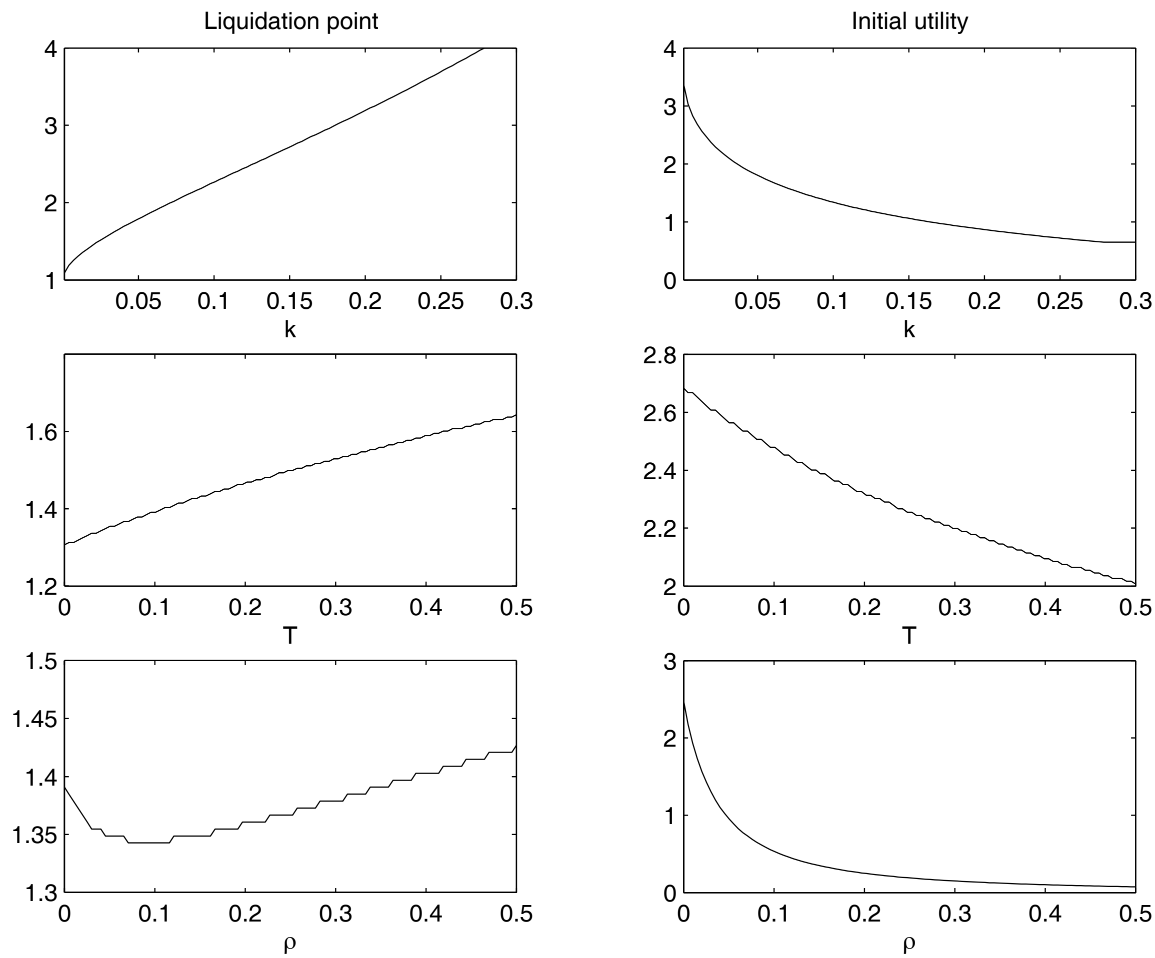

Figure 3. The graphs show, for an investor who derives utility from realized gains and losses, how the liquidation point at which he sells a stock and the initial utility from buying it depend on the transaction cost k, the length of the search interval T between the sale of one stock and the purchase of another, and the arrival rate of an exogeneous liquidity shock ρ.

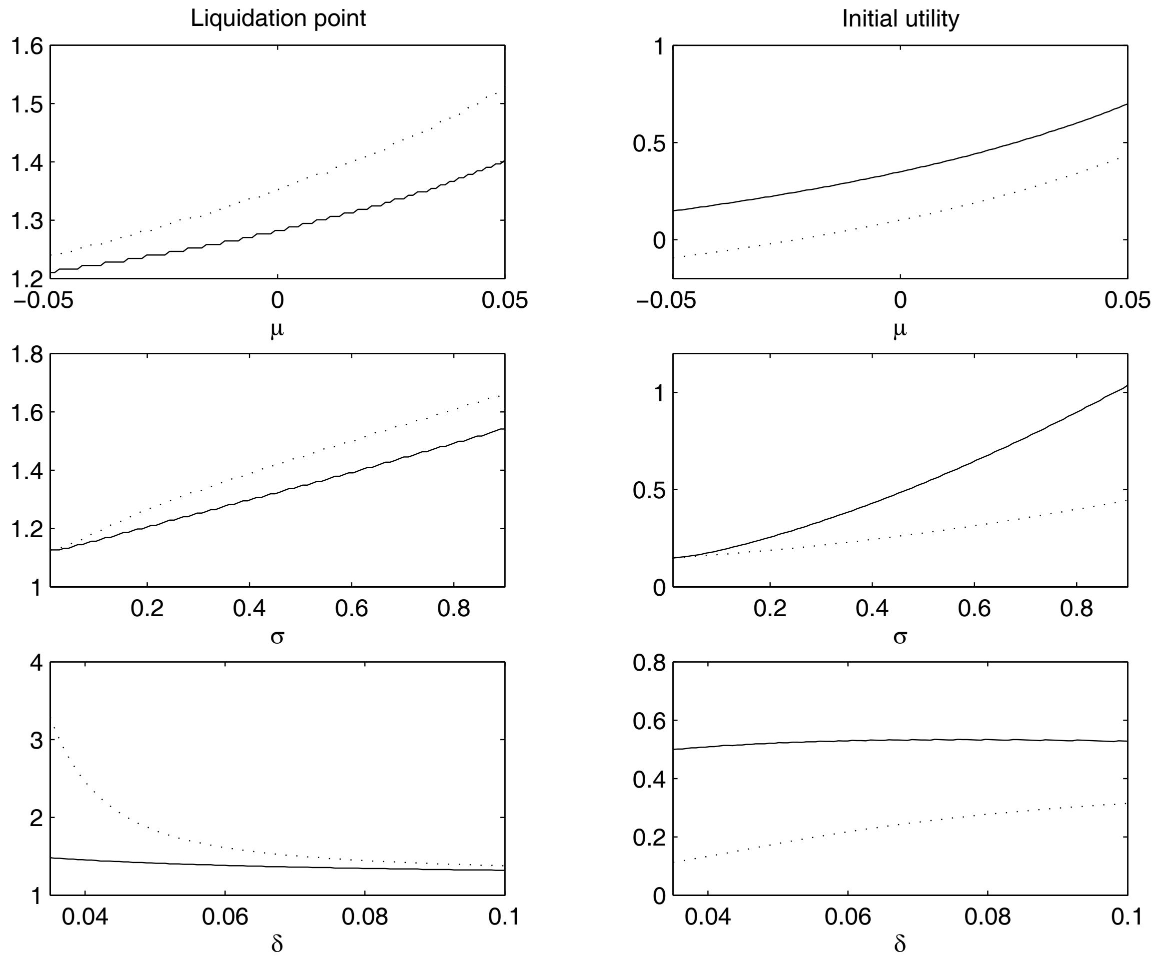

Figure 4. The graphs show, for an investor who derives utility from realized gains and losses, how the liquidation point at which he sells a stock and the initial utility from buying it depend on the stock's expected return μ, its standard deviation σ, and the investor's discount rate δ. The solid lines correspond to a model that allows for a transaction cost, a search interval between the sale of one stock and the purchase of another, and an exogeneous liquidity shock. The dotted lines correspond to a model which also includes these features, but in which the functional form of realization utility is piecewise linear, rather than linear.

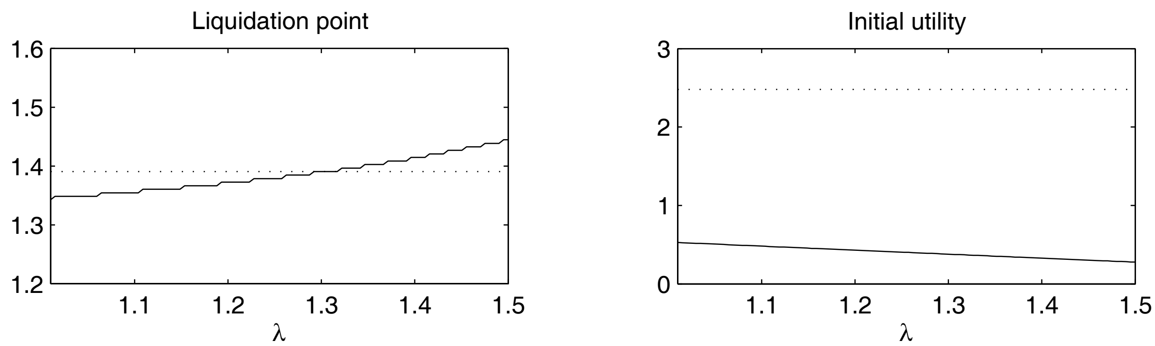

Figure 5. The graphs show, for an investor who derives utility from realized gains and losses, how the liquidation point at which he sells a stock and the initial utility from buying it depend on the investor's degree of loss aversion λ. The solid lines correspond to a model with liquidity shocks, and the dotted lines to a model with no liquidity shock.

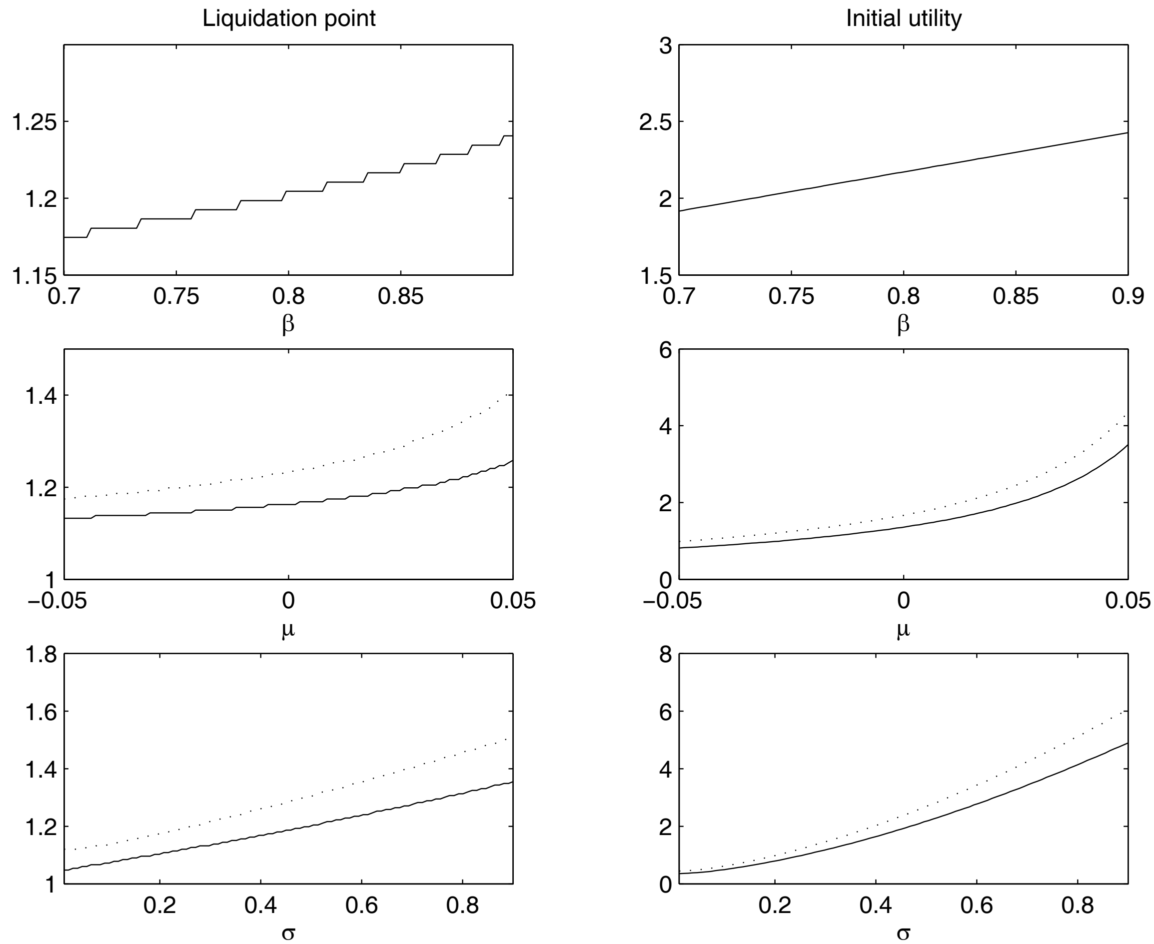

Figure 6. The solid lines in the graphs show, for an investor who derives utility from realized gains and losses and exhibits hyperbolic time discounting, how the liquidation point at which he sells a stock and the initial utility from buying it depend on the hyperbolic discounting parameter β, the stock's expected return μ, and its standard deviation σ. The dotted lines in the middle and bottom panels correspond to a model with standard exponential time discounting.

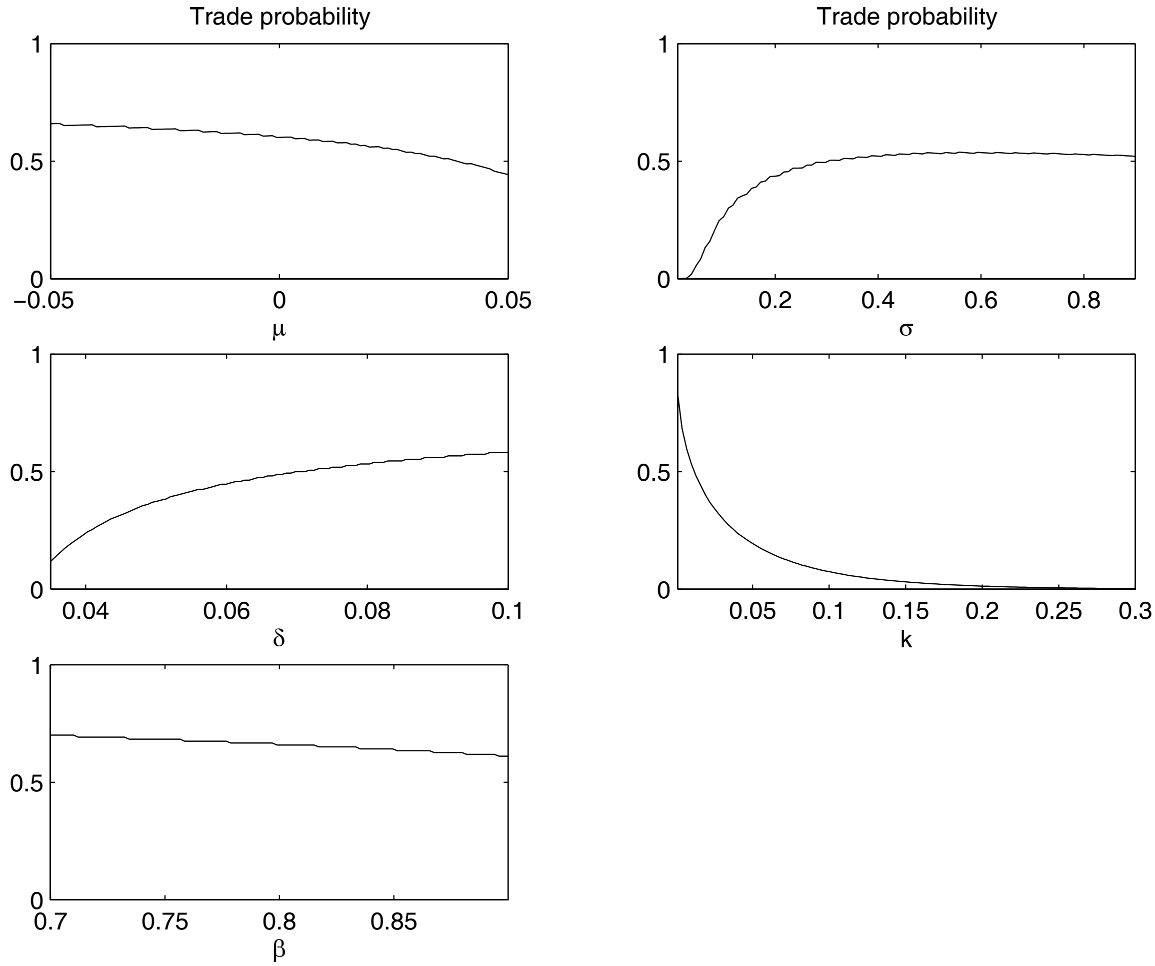

Figure 7. The graphs show, for an investor who derives utility from realized gains and losses, how the probability that the investor will sell a specific stock within a year of buying it depends on the stock's expected return μ, its standard deviation σ, the exponential time discount rate δ, the transaction cost k, and the hyperbolic time discount rate β.

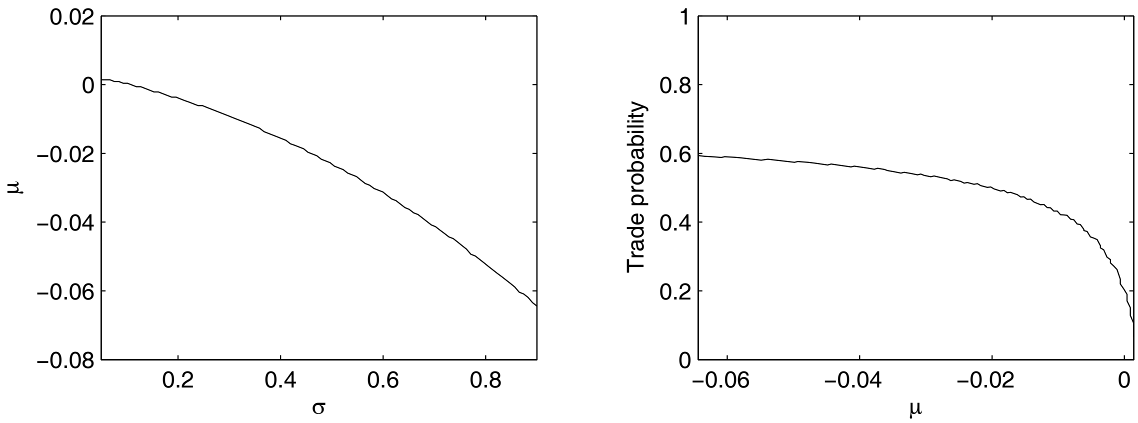

Figure 8. The top-left graph shows, for an economy populated by investors who derive utility from realized gains and losses, the equilibrium relationship between expected return and standard deviation in a cross-section of stocks. The top-right graph shows, for the same cross-section, the equilibrium relationship between expected return and trading intensity.