

Realization Utility
with Reference-Dependent Preferences[†]

Jonathan E. Ingersoll, Jr.

Yale School of Management

Lawrence J. Jin

Yale School of Management

July 22, 2012

[†]We thank our colleagues at the Yale School of Management and the seminar participants at the LBS Trans-Atlantic Doctoral Conference, the University of Massachusetts, Amherst, and the 2012 Western Finance Association Annual Meeting for helpful discussions. In particular we would like to thank Nick Barberis, David Hirshleifer, and an anonymous referee. Jin acknowledges support from a Whitebox Advisors grant.

Realization Utility with Reference-Dependent Preferences

Jonathan E. Ingersoll, Jr. and Lawrence J. Jin

Yale School of Management

Realization utility, the concept that investors derive their utility only from actively realizing gains or losses, has drawn increasing interest recently. In this paper we develop a tractable model of realization utility that significantly extends existing models both psychologically and institutionally. We study the roles of S-shaped Prospect Theory utility functions in a dynamic context and show that our model generates voluntary sales at both gains and losses. The model can be calibrated to a variety of statistics observed in the data, including Odean's measure of the disposition effect. It makes specific predictions about the relative volume of gains and losses, the duration of holding periods, and the sizes of both realized and paper gains and losses. In addition, our model yields predictions such as the flattening of the capital market line and the pricing of idiosyncratic risk.

JEL classification: G02, G11, G12

Keywords: Realization Utility, Prospect Theory, Disposition Effect

Realization Utility with Reference-Dependent Preferences

1. Introduction

How do investors decide on and evaluate their own investment performance? Standard economic theory posits that investors maximize the expected utility of their lifetime consumption stream by dynamically adjusting their portfolio allocations based on their current wealth and their expectations of the future. Although this way of modeling investors' behavior may be close to reality for sophisticated investors such as mutual fund managers, it is questionable whether less sophisticated investors behave this way.

A growing amount of research shows that individual investors do not always behave in the ways that expected utility theory predicts. In particular, the independence axiom seems troublesome as does the assumption of risk aversion, at least for losses. The latter assumption is not a requirement of expected utility theory, but it or something like it is important for most equilibrium models which follow from maximizing behavior. In contrast to what theories like the APT or CAPM predict, individual investors seem to be particularly concerned about an asset's change in price from a reference point.

Behavioral literature suggests alternative views of modeling investors' behavior. Kahneman and Tversky (1979) have argued that risk aversion does not characterize many choices and proposed an S-shaped utility function. Shefrin and Statman (1985) use "mental accounting" to justify investors' concentrating on specific separate incidents. Thaler (1999) says "A realized loss is more painful than a paper loss." Barberis and Xiong (2012) study a model which assumes that investors think of their investing experience as a series of separate episodes during each of which they either made or lost money and that the primary source of utility comes in a burst when a gain or loss is realized.¹ Frydman et al. (2012) find evidence using the neural data that supports this "realization utility" hypothesis.

In this paper, we use these notions to develop an intertemporal model of investors who have Prospect Theory's S-shaped utility and who evaluate their performance incident by incident based on realized profits and losses.² Our model is a partial equilibrium framework with an infinite horizon. An investor purchases stocks whose prices evolve as geometric Brownian motion. At each subsequent point in time, the investor decides whether to hold onto his current investment or realize his gain or loss thereby obtaining an immediate utility burst. If he sells, he reinvests the proceeds after transaction cost into another stock. We show that the investor's optimal strategy is to wait until the stock price rises or falls to certain percentages above or below the purchase price before selling. Our model includes that of Barberis and Xiong (2012) as a special case.

¹ Further discussion on the psychological foundation of viewing investments as episodes is in their paper.

² Kyle, Ou-Yang, and Xiong (2006) study a one-time liquidation problem with prospect theory preferences. But reinvestment, which is a key component of our model, is ignored in their model.

One important difference between these two models is that ours adopts an S-shaped utility while theirs uses piecewise linear utility. As a direct consequence, they predict that investors voluntarily realize gains but never voluntarily sell at a loss, which is clearly unrealistic and inconsistent with the data. With an S-shaped function, marginal utility decreases with the magnitude of both gains and losses. This means that lifetime utility can be increased by taking frequent small gains along with occasional larger losses, because the latter have less total disutility than the utility of the former, and realizing losses resets the reference level for and future gains. The disposition effect, the tendency to realize gains but defer losses, follows naturally from this result but it is a *dynamic* result; voluntary loss taking is optimal only because the subsequent reinvestment resets the reference level and increases the likelihood of realizing future gains.

What has been commonly argued in both theoretical and empirical literatures is that an S-shaped utility function leads to the disposition effect because risk seeking over losses induces investors to retain their positions and gamble on the future while risk aversion over gains induces the opposite. However, this is a *static* argument. Extrapolating this reasoning period by period would imply losses are never realized; the disposition effect should be infinite. In our dynamic realization utility model exactly the opposite is true. Investors naturally want gains, but an S-shaped utility helps to generate voluntary losses and thereby reduces the magnitude of the disposition effect to the observed level.

We calibrate our model in two parts. First we show that the magnitudes and frequencies of realized gains and losses and the frequencies of paper gains and losses as observed in the trading data of Odean and others are consistent with the type of simple two-point strategy our model predicts. In particular, Odean reports average gain and loss sizes of 28% and -23% with 54% of trades being realized gains and that when trading investors realize 15% of possible gains and 10% of possible losses. We show that using those average realized gain and loss sizes, our model predicts that 58% of sales should be gains, and investors should have a propensity of 14% and 11% to realize gains and losses, respectively. In addition, we propose a modified form of Tversky-Kahneman utility that can generate these two sales points, 28% and -23% , either alone or in a mixture of heterogeneous investors.

Our model also has a variety of other empirical implications and predictions. For instance, investors may be risk-seeking in some circumstances due to the option value inherent in realizing losses; this helps explain a flatter security market line and the negative pricing of idiosyncratic risk as shown in Ang, Hodrick, Xing, Zhang (2006). It may also help explain why investors appear to hold portfolios that appear under-diversified.

The plan of our paper is as follows. In Section 2, we lay out a specific intertemporal reference-dependent realization utility model and present its solution and basic insights. Section 3 examines the properties of the derived value function and analyzes the optimal sales policies. Section 4 provides a detailed calibration of our model to several empirical regularities. Section 5 analyzes voluntary loss realization at a general level. Section 6 presents further model applications and predictions. Section 7 gives some concluding remarks and direction for future research. A summary of the important notation, all the proofs, and some more detailed technical considerations are included in the Appendix.

2. Realization Utility Model with Tversky-Kahneman Utility

In this section we present a simple, specific model of intertemporal realization utility. Our investor takes positions in a series of stocks, buying a number of shares and later selling his entire position and reinvesting it. Each realized gain or loss contributes a burst of utility, and our investor acts as to maximize the expectation of the sum of the discounted values of these bursts.

We assume the investor applies narrow framing when he evaluates his gains. This assumption side-steps any complications that might arise from diversification or rebalancing motives.³ Narrow framing can be justified if the investor derives realization utility only when both the purchase and sale prices of the asset are salient, and, therefore, evaluating *individual* assets is the applicable setting for studying realization utility. As a result, even when the investor holds multiple stock positions simultaneously, narrow framing allows us to study each sequence of purchases and sales separately.

Secondly, we assume that a utility burst is received only at the time when a gain or a loss is realized. As with Prospect Theory, we normalize utility so that gains and losses contribute positive and negative utility, respectively.⁴ While it is assumed utility depends primarily on the size of the gain or loss, G , it seems reasonable that the reference level, R , might also have a separate effect. In particular, a gain or loss of a given size probably has a greater utility impact, either good or bad, the smaller is the reference level; e.g., the gain or loss of \$10 is felt more strongly when the reference level is \$100 than when it is \$500. Therefore, we denote the utility burst function as a function of both variables, $U(G, R)$. In this paper we assume that $U(G, R)$ is homogeneous of degree β in G and R

$$U(G, R) = R^\beta u(G/R). \quad (1)$$

This assumption is important for keeping the model tractable, but it also focuses utility on the rate of return rather than dollar changes which is in keeping with the general emphasis in Finance. Expressed in this way, the scaling parameter β gauges the impact of the reference level on utility bursts measured as a rate of return.

We study a Merton-type partial equilibrium economy in continuous time with infinite horizon. At $t = 0$, the investor chooses either to stay out of the market which earns him a normalized utility of zero or to invest in one of a number of identically distributed stocks. The stock prices evolve according to a geometric Brownian motion $dS/S = \mu dt + \sigma d\omega$ where μ and σ are the growth rate and logarithmic standard deviation, respectively, and ω is a standard Brownian motion. At each subsequent point in time, t , the investor chooses to either hold his investment for

³ For instance, the investor might have incentive to sell a losing stock to purchase a winning stock. Another example could be the incentive of purchasing diversified fund. These considerations are outside the scope of this paper though some related discussion is provided in the last two sections of the paper.

⁴ Aside from its simple intuition, centering $U(0, R) = 0$ makes the magnitude effect discussed next easy to characterize. Typically the centering of utility is arbitrary and has no effect on expected utility maximization. In some models like this, the investor might be able to choose to take no action at all so if no action is presumed to give a utility of zero, then the centering chosen here can affect participation.

longer time or sell his entire position and realize a utility burst. When he sells, he pays a proportional transaction cost, k_s , and reinvests the net proceeds after paying a second proportional transaction cost, k_p , reducing his investment to $X_{t^+} = (1 - k_s)X_{t^-} / (1 + k_p) \equiv KX_{t^-}$. Between realization dates, the investment value follows the same geometric Brownian motion as the underlying asset

$$dX/X = \mu dt + \sigma d\omega. \quad (2)$$

In a static Prospect Theory setting, the reference level is essentially a parameter of the utility function defining the status quo. However, in our dynamic model, we must address how it is updated and exactly how the gain or loss is measured relative to it. The simplest rule is R is set at the net purchase price, as defined above, and remains constant between sales.⁵ That is, when the investor sells his stock for X_t , he resets his reference level to KX_t until the next sale. However, this is a subjective matter and could differ from investor to investor; there are other ways that the new reference level might be set. For instance, an investor might view R as the gross amount invested including the purchasing cost, $(1 - k_s)X_t$. It might also be some intermediate level particularly if the transaction cost has different components such as a bid-ask spread and a commission. Most brokerage accounts show the purchase price which would tend to emphasize the net investment as the reference level. On the other hand, the taxable capital gain excludes the purchase cost which would tend to emphasize the gross investment as the reference level. In our analysis, we assume the simplest case that investor fully accounts for costs and sets the reference level to the net amount invested, KX .⁶

A related issue is how the investor evaluates his utility burst upon a sale. Again, there are several ways he might do so depending on his subjective view of the transaction costs. For example, if he ignores costs completely, then the gain is the gross sales value less the reference level, $G_t = X_t - R_t$. If he fully recognizes transaction costs and compares the net reinvested amount to the reference level, then $G_t = KX_t - R_t$. If he views the gain as the difference between the net proceeds of the sale and the reference level, then $G_t = (1 - k_s)X_t - R_t$. These three cases are covered by defining the gain as $G_t = \kappa X_t - R_t$, and setting the parameter κ to 1, K , or $1 - k_s$, respectively. Intermediate views are also possible. We leave the parameter κ free allowing many interpretations.

The time-consistency of these rules together with the assumptions that (i) the utility bursts in (1) are homogeneous of degree β in X and R , (ii) the asset value process has stochastic constant returns to scale, and (iii) the investment horizon is infinite jointly guarantee that the future looks the same depending only on the current investment and reference level. This

⁵ In some setting, the reference level might grow deterministically at a constant rate (for instance, the interest rate), or evolve stochastically over time. It could also be updated based on recent history of the stock price. These considerations are omitted in this paper for length limit.

⁶ The analysis here is largely unchanged for different ways of setting the reference level. If an investor adopts the gross cost view, then equation (5) below has $V(KX, (1 - k_s)X)$ as the second term on the right-hand side. Presumably, an investor would not adopt a gross cost view for setting the new reference level as well as recognizing both costs in assessing the gain as this would double count the purchasing costs. However, the only requirement for our model is that the investor sets his reference level *consistently* over time. Barberis and Xiong (2012) also adopt the net cost interpretation which in their notation is $(1 - k)X$ with k being the round-trip transaction cost and only consider the full recognition of transaction costs in determining gain size.

simplifies our problem in two ways. First, it removes time as an explicit variable. Second, our investor always has the incentive to reinvest immediately upon selling a position provided he chose to enter the market in the first place.

Denote the undiscounted value function at time t by $V(X_t, R_t)$. As discussed above, V does not depend on time explicitly but only the current investment and reference level. By definition, the value function is the maximized expectation of the sum of future discounted utility bursts

$$V(X_t, R_t) = \max_{\{\tilde{\tau}_i\}} \mathbb{E}_t \left[\sum_i e^{-\delta \tilde{\tau}_i} U(\tilde{G}_{t+\tilde{\tau}_i}, \tilde{R}_{t+\tilde{\tau}_i}) \right] \quad (3)$$

where δ is the rate of time preference, $\tilde{G}_{t+\tilde{\tau}_i}$ and $\tilde{R}_{t+\tilde{\tau}_i}$ are the dollar size and the reference level for the i^{th} future gain, respectively, and $\tilde{\tau}_i$ is the random time at which it is realized. In our model, these are stopping times that are endogenously chosen by the investor to maximize his lifetime expected utility. To solve the problem posed by (3), we use the time-homogeneity property to rewrite it as a recursive expression

$$V(X_t, R_t) = \max_{\tilde{\tau}} \mathbb{E}_t \left[e^{-\delta \tilde{\tau}} U(\kappa X_{t+\tilde{\tau}} - R_t, R_t) + e^{-\delta \tilde{\tau}} V(KX_{t+\tilde{\tau}}, KX_{t+\tilde{\tau}}) \right] \quad (4)$$

where $\tilde{\tau}$ is the next sales time. Hereafter, we suppress time subscripts for notational convenience unless necessary for clarity.

At a sale, the value function before the sale must equal the sum of the utility burst of the sale and the post-reinvestment continuation value function. So upon a sale realizing X before costs,

$$V(X, R) = U(\kappa X - R, R) + V(KX, KX). \quad (5)$$

Between sales times, equation (4) can be re-expressed using the law of iterated expectations and Ito's lemma

$$0 = \mathbb{E}\{d[e^{-\delta t} V(X, R)]\} = e^{-\delta t} \left(\frac{1}{2} \sigma^2 X^2 V_{XX} + \mu X V_X - \delta V \right) dt. \quad (6)$$

Because $U(G, R)$ is homogeneous of degree β in G and R and the asset value process has stochastic constant return to scale, V must be also homogeneous of degree β in X and R and therefore can be written as $V(X, R) = R^\beta v(x)$ where $x \equiv X/R$ is the one plus the rate of return measured on the reference value.⁷ The equation for this reduced form value is

$$0 = \frac{1}{2} \sigma^2 x^2 v'' + \mu x v' - \delta v. \quad (7)$$

The general solution to (7) is

⁷ V must be positively homogeneous; i.e., $\beta \geq 0$. Since $V(X_0, X_0) = X_0^\beta v(1)$, utility is decreasing in the amount of the original investment when $\beta < 0$, and the investor would always prefer to reduce his initial investment. A positive β also assures that $|U(gR, R)|$ is increasing in R for a fixed g ; that is, the higher the reference level, the bigger is the utility of a given rate of return. This property is similar to increasing relative risk aversion.

$$v(x) = C_1 x^{\gamma_1} + C_2 x^{\gamma_2} \quad \text{where} \quad \gamma_{1,2} \equiv \frac{-\mu + \frac{1}{2}\sigma^2 \pm \sqrt{(\mu - \frac{1}{2}\sigma^2)^2 + 2\delta\sigma^2}}{\sigma^2}. \quad (8)$$

This is true regardless of the form for u . The utility of the sales bursts affects only the constants, C_1 and C_2 . Again due to the homogeneity, the optimal sale strategy must be to realize a gain or loss when the stock price reaches a constant multiple, Θ , or fraction, θ , of the reference level.⁸ The upper sale point, Θ , must exceed $1/\kappa > 1$ as otherwise the sale is not a gain after costs. The lower sale point, θ , must be less than 1.⁹

Applying the homogeneity relation (1) to the boundary condition (5) yields the reduced-form boundary conditions

$$v(\Theta) = u(\kappa\Theta - 1) + (K\Theta)^\beta v(1), \quad v(\theta) = u(\kappa\theta - 1) + (K\theta)^\beta v(1). \quad (9)$$

Equating these to the general solution from (8), we can determine the constants C_1 and C_2 in terms of the policy variables

$$C_1 = \frac{c_2(\theta)u(\kappa\Theta - 1) - c_2(\Theta)u(\kappa\theta - 1)}{c_1(\Theta)c_2(\theta) - c_1(\theta)c_2(\Theta)} \quad (10)$$

$$C_2 = \frac{c_1(\Theta)u(\kappa\theta - 1) - c_1(\theta)u(\kappa\Theta - 1)}{c_1(\Theta)c_2(\theta) - c_1(\theta)c_2(\Theta)} \quad \text{where} \quad c_i(\phi) \equiv \phi^{\gamma_i} - (K\phi)^\beta.$$

The optimal sales points, Θ and θ , can now be determined either by maximizing $C_1 + C_2$, which is the value of $v(1)$, or by applying the smooth-pasting condition at both sales points.¹⁰ The solution and the optimal θ - Θ strategy are illustrated in Figure 1. The no-sales region runs from θ to Θ . The value function exceeds the sum of the utility burst plus the continuation value in this region as illustrated; it is tangent to the payoff including continuation value at Θ and θ . For some parameter values, it may be optimal to forgo all losses. In these cases, the continuation value is not large enough to offset that disutility of realizing a loss. The value function is still given by (8) though $C_2 = 0$ and, therefore, $v(0) = 0$.

⁸ See the Appendix for more details on the constancy of the optimal policy.

⁹ A sale at any point in the range $(1, 1/\kappa)$ produces a subjective loss after accounting for transaction costs so such a sale could never be part of an optimal strategy. Under a constant policy with θ in this range, there would never be any sales at a higher price as the stochastic process for x is continuous and begins at 1 after each repurchase when the reference level is set to the net investment. But this means that only losses with their negative utility bursts would be realized leading to a negative v . This could not be the optimal policy as never selling gives a utility of zero as does not participating at all.

¹⁰ The optimal sales strategy must maximize v for every value of its argument in the continuation region, and $x = 1$ is guaranteed to be in the continuation region since $\theta < 1 < 1/\kappa < \Theta$. Note that the smooth pasting condition does not simply match the derivative of v to the marginal utility of the burst. It must be applied to (9) which has the continuation value as well as the utility burst on the right-hand side. In some cases there is a constrained optimum, $\theta = 0$, at which the smooth pasting condition does not apply as discussed in Proposition 1 below. Certain parameter restrictions must hold to solve this problem. Unless $\delta > 0$ and $\beta < \gamma_1$, there is no unique optimum as many strategies lead to infinite utility. These transversality issues are discussed in the Appendix.

Value Function and Optimal Sales Policies illustrating optimal sales at gains and losses

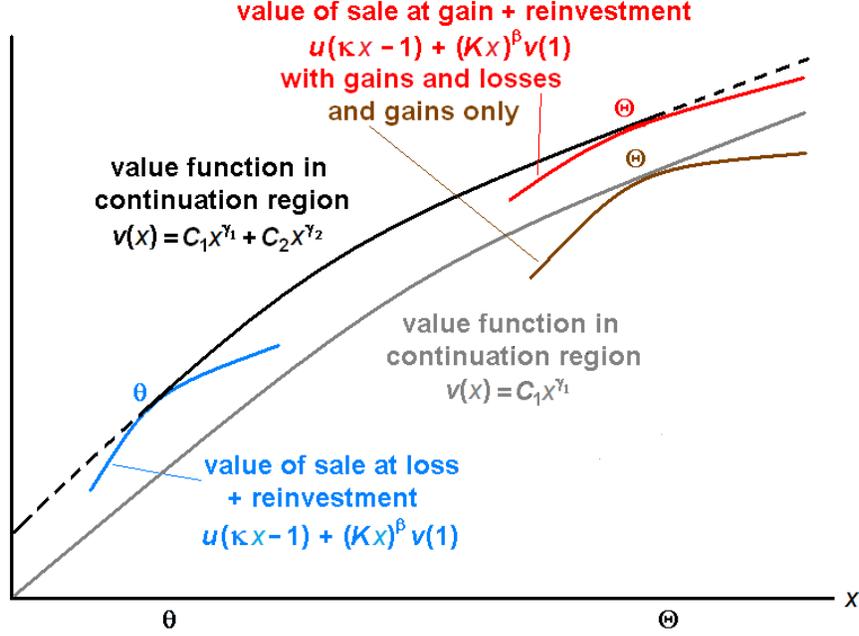


Figure 1: Determination of the Optimal Sales Policies. This figure illustrates the value function and the optimal policy for realizing gains and losses. The value in the continuation or no-sales region is tangent to the payoff functions at both sales points, θ and Θ . In this illustrated case, it is optimal to realize some losses.

Consider typical stock and transaction costs values: $\mu = 9\%$, $\sigma = 30\%$, and $k_s = k_p = 1\%$. How would a realization-utility investor trade this stock? To answer this question, we must specify the burst utility function. A reference-scaled version of the Cumulative Prospect Theory (CPT) utility proposed by Tversky and Kahneman (1992), hereafter called scaled-TK utility, is

$$U_{sTK}(G, R) = R^\beta u_{sTK}(G/R) \quad \text{for} \quad u_{sTK}(g) = \begin{cases} g^{\alpha_G} & g \geq 0 \\ -\lambda(-g)^{\alpha_L} & g < 0 \end{cases} \quad (11)$$

with $0 < \alpha_G, \alpha_L \leq 1$, $\lambda \geq 1$.¹¹ As for CPT, the parameters α_G and α_L determine the investor's risk aversion over gains and risk seeking over losses while loss aversion is measured by λ . The scaling parameter satisfies $0 \leq \beta \leq \min(\alpha_L, \alpha_G)$. As discussed earlier, the upper restrictions on β ensures that $|U(G, R)|$ is decreasing in R for a fixed G and a nonnegative β is a participation constraint.

¹¹ Setting $\beta = \alpha_G$ and $\lambda = \lambda_0 R^{\alpha_L - \alpha_G}$ reduces (11) to the standard case introduced in Tversky and Kahneman (1992), with λ_0 being their loss aversion parameter. Loss aversion then would vary with R , but the reference level is constant in the original static interpretation. The Barberis and Xiong (2012) model is the special case $\beta = \alpha_G = \alpha_L = 1$.

Tversky and Kahneman (1992) estimated utility parameters, $\alpha_G = \alpha_L = 0.88$ and $\lambda = 2.25$. Since they were not concerned about intertemporal aspects, they did not estimate a discount rate nor did they consider scaling. However, for such a low level of risk aversion, the transversality condition is violated and the investor waits forever to realize any gain unless δ is nearly equal to the expected rate of return.¹² For $\alpha_G = \alpha_L = 0.88$ and $\delta = 8\%$, the investor never voluntarily realizes losses unless there is little loss aversion with λ close to one.

However, voluntary loss taking can be part of the optimal policy for other utility parameters. Wu and Gonzalez (1996), for example, estimate $\alpha = 0.5$. Using utility parameters, $\alpha_G = \alpha_L = 0.5$, $\lambda = 2$, and $\delta = 5\%$ the optimal strategy does include voluntary losses for any β less than about 0.327. Figure 2 shows the optimal sales strategy for several parameter sets plotting Θ and θ against β .¹³ Provided any losses are realized, they are always larger than gains in magnitude. This might seem counterintuitive, but these smaller gains are realized more frequently, and since marginal utility is decreasing with the magnitude of the gain or loss, several small gains more than offset the disutility of a single loss of the same total size.

One commonly discussed observation about the realization of gains and losses is the disposition effect, which has often been claimed to be a consequence of an S-shape utility function.¹⁴ The reasoning goes that the investor realizes sure gains as he is unwilling to gamble about future gains; however, being risk-seeking over losses he will gamble and postpone realizing sure losses. This reasoning implicitly assumes something like realization utility because any effects of unrealized gains or losses are ignored, but it is not truly dynamic as it ignores any effects of reinvestment after sales. Nor does it address the question of why any losses are ever realized and not just continually postponed.

With concave realization utility, the disutility of a loss can never be offset by the benefit of recovering that loss in subsequent gains of the same total size, but as seen in Figure 2 losses will be realized as well as gains with an S-shaped utility function. For the parameters used that losses are substantially less common than gains since θ is much farther from 1 than is Θ .¹⁵ However, in sharp contrast with the static argument mentioned above that an S-shaped utility leads to the disposition effect, the S-shape actually serves to reduce the disposition effect in a dynamic context. As the S-shape becomes more pronounced with both α_G and α_L decreasing, the

¹² The restriction on δ comes from the transversality condition, $\alpha_G \leq \gamma_1$. While the required discount rate is large relative to those usually assumed, many behavioral finance models do assume that investors are quite impatient. It seems reasonable that utility derived from trading gains might well display more impatience than utility for lifetime consumption. In addition, this high discount rate could incorporate the hazard rate describing the investor's ceasing this type of trading. Death is sometimes inserted into infinite-horizon models in this fashion, though here the termination of trading might be a simple lack of further interest.

¹³ As discussed in footnote 10, $\beta < \gamma_1$ ensures that a unique optimal policy exists. As $\beta \rightarrow \gamma_1$, $\Theta \rightarrow \gamma_1 / [\kappa(\gamma_1 - \alpha_G)]$.

¹⁴ Shefrin and Statman (1985) were the first to argue that an S-shaped utility function lead to the disposition effect. Similar arguments were made in Weber and Camerer (1998), Odean (1998), Grinblatt and Han (2005), and other theoretical and empirical papers.

¹⁵ Since returns are lognormal the proper "distance" comparison is $\ln \Theta \ll -\ln \theta$; also $\mu - \frac{1}{2}\sigma^2 > 0$ so even if the log distances were equal, gains would be realized more often.

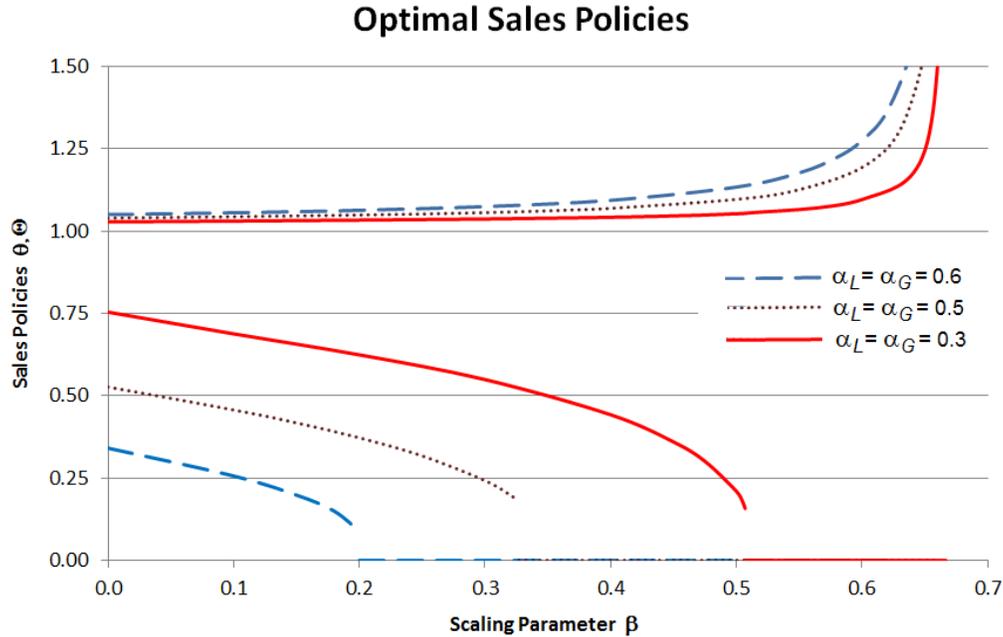


Figure 2: Optimal Sales Policies. This figure illustrates the optimal policies for selling at gains or losses as a multiple or fraction of the reference level. The stock price parameters are $\mu = 9\%$, $\sigma = 30\%$. Transaction costs are $k_s = k_p = 1\%$. Utility parameters are $\delta = 5\%$, $\lambda = 2$, and $\alpha_G = \alpha_L = 0.3, 0.5, 0.6$, as indicated. The investor fully recognizes transaction costs in assessing his utility, i.e., $\kappa = K = (1 - k_s)/(1 + k_p)$. The scaling parameter β must be less than α_G if utility is to decrease in magnitude as the reference level rises and must be less than γ_1 ($= 2/3$ for these parameters) for a unique Θ to exist. As $\beta \rightarrow \gamma_1$, $\Theta \rightarrow \gamma_1/[\kappa(\gamma_1 - \alpha_G)]$.

optimal gain point is affected only a little while the optimal loss point increases dramatically substantially reducing the disposition effect. The reason for this is that realizing a loss resets the reference level for future possible gains, and this can more than offset the direct disutility of the loss. That is, the realization of a loss is, in some sense, the purchase of a valuable option. When α_G is small, this effect can be substantial since the marginal utility of small gains is very large inducing the investor to repeatedly realize small gains.

With intertemporal realization utility, an S-shaped utility function does not create the disposition effect; it actually reduces it explaining why any voluntary losses are realized at all rather than none.¹⁶

The scaling parameter, β , also plays an important role in the optimal policies. When $\beta < \alpha_G$ as assumed, gains earned on a smaller reference level are valued more highly. So when β is small, the pain of realizing a loss is offset more by the lowering of the reference level for subsequent gains. This intensifies loss taking (increases θ). It also strengthens gain taking but cannot

¹⁶ Some research suggests that Prospect Theory may not lead to the disposition effect. See for example, Barberis and Xiong (2009), Kaustia (2010), and Hens and Vlcek (2011), but none of these papers consider reinvestment.

have a large effect on Θ which is already quite near 1. On the other hand, as seen in Figure 2, as β becomes large, the gain sales point, Θ , increases while the loss sales point, θ , falls. As shown in (9), after a sale, the continuation payoff is proportional to $(K\Theta)^\beta$ or $(K\theta)^\beta$. Therefore, a very high β makes realizing large gains quite beneficial by substantially increasing the continuation payoff. On the other hand, losses are much less valuable and θ drops discontinuously to zero.

The discontinuous change in θ occurs because this maximization problem is not a standard convex optimization. As illustrated in Figure 3, the reduced value function, $v(1)$, is not a concave function of θ . Both an interior local maximum and a corner local maximum at zero are possible and either can be the global maximum. The high marginal disutility of repeated small losses together with transaction costs makes loss taking suboptimal for high values of θ near 1. On the other hand for low θ , the continuation value, which is proportional to $(K\theta)^\beta$, is very small and cannot offset the disutility of a loss. This makes avoiding losses altogether ($\theta = 0$) better than taking a large loss. Only for intermediate values of θ is the continuation value possibly sufficient to offset the disutility of a loss. So $v(1)$ attains its local maximum value either at $\theta = 0$ or an intermediate value.

Figure 3 illustrates both optimum types for $\alpha_G = \alpha_L = 0.5$, $\beta = 0.3$. The initial value function after any sale and repurchase, $v(1)$, is plotted against θ for three values of λ . The upper sales point is fixed at its distinct optimal value in each case. For $\lambda = 2.5$, it is optimal to sell for a loss at $\theta = 0.183$. For $\lambda = 2.56$, there is an optimum at $\theta = 0.147$, but this is only a local maximum as never selling at a loss provides higher utility as shown. For the critical value of $\lambda \approx 2.531$, both selling for a loss at $\theta = 0.166$ and never selling for a loss provide the same expected utility. Therefore, the lower sales point, θ , does not decrease smoothly to zero as λ increases; it drops discontinuously from 0.166 to 0 as λ passes the critical value of 2.531. A similar change in regime is true for the other parameters. The two regimes are characterized in Proposition 1; a proof is supplied in the Appendix.

Proposition 1: Scaled-TK utility has both an upper and a (non-zero) lower optimal sales point if and only if λ is less than the critical value

$$\lambda_* = \frac{(\kappa\Theta_* - 1)^{\alpha_G - 1} \Theta_*^\beta}{(1 - \kappa\theta_*)^{\alpha_L - 1} \Theta_*^\beta} \times \frac{(\alpha_G - \gamma_1)\kappa\Theta_* + \gamma_1}{(\alpha_L - \gamma_1)\kappa\theta_* + \gamma_1}$$

where

$$\begin{aligned} 0 &= (\alpha_G - \gamma_1)\kappa\Theta_*^{\gamma_1 + 1 - \beta} + \gamma_1\Theta_*^{\gamma_1 - \beta} - (\alpha_G - \beta)K^\beta\kappa\Theta_* - \beta K^\beta \\ 0 &= (\alpha_L - \gamma_1)\kappa\theta_*^{\gamma_1 + 1 - \beta} + \gamma_1\theta_*^{\gamma_1 - \beta} - (\alpha_L - \beta)K^\beta\kappa\theta_* - \beta K^\beta \end{aligned}$$

(12)

determine Θ_* and θ_* . If λ is greater than this critical value, only gains are realized. The solution is still characterized (9) and (10) with C_2 set to 0. As β approaches its transversality upper limit, γ_1 , $\lambda_* \rightarrow 0$, and voluntary losses are never realized. ■

The Barberis-Xiong (2012) model is a special case of scaled-TK utility with $\alpha_L = \alpha_G = \beta = 1$. For this model, or indeed any realization utility model with piece-wise linear utility for gains

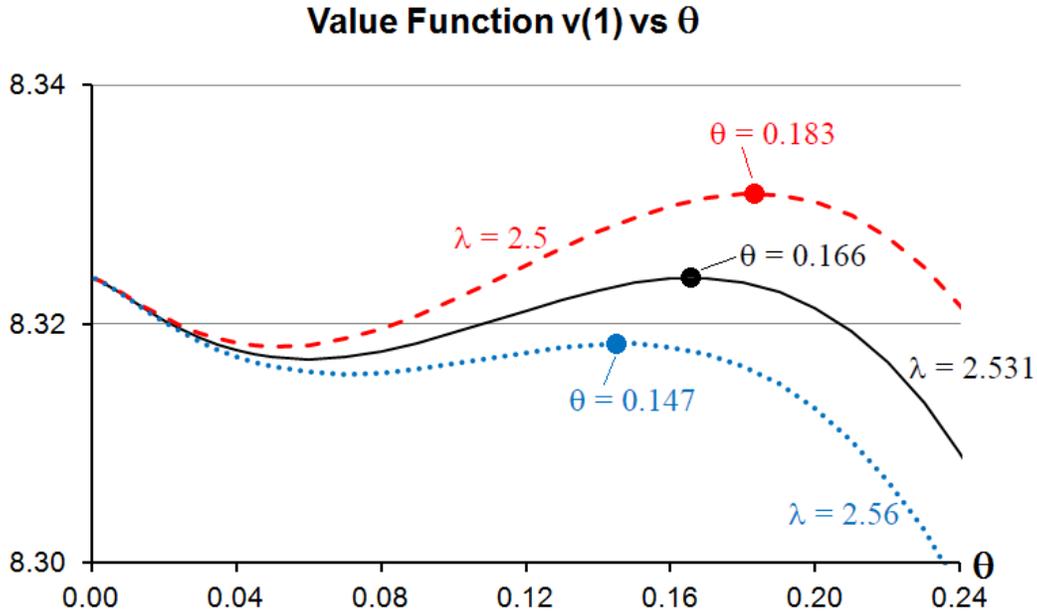


Figure 3: The Value Function for Scaled-TK Utility. The reduced value function for scaled-TK utility measured immediately after a sale and reference level reset, $v(1)$, is plotted against different loss sales points, θ . The gain sales point, Θ , is fixed at its optimal value. The solid black line shows the value function for $\lambda = 2.531$, the dashed red line shows the value function for $\lambda = 2.5$, and the dotted blue line shows the value function for $\lambda = 2.56$. The other parameters are $\mu = 9\%$, $\sigma = 30\%$, $k_s = k_p = 1\%$, $\alpha_G = \alpha_L = 0.5$, $\beta = 0.3$, $\delta = 5\%$ $\kappa = 1 - k_s$. For $\lambda = 2.5$, the two-point policy ($\theta = 0.183$, $\Theta = 1.037$) is optimal. For $\lambda = 2.56$, the one-point policy ($\Theta = 1.036$) is optimal. For the critical value $\lambda = \lambda_* = 2.531$, the two point policy ($\theta = 0.166$, $\Theta = 1.036$) and the one-point policy ($\Theta = 1.036$) have the same expected utility.

and losses and $0 \leq \beta < \gamma_1$, the critical value λ_* is less than 1. Therefore, losses are never realized voluntarily.

3. The Value Function and Optimal Sales Policies

The value function or its reduced-form equivalent, v , measures the present value of the investor's utility bursts and gives a point estimate of the benefit of his strategy. It serves the role of the derived utility function in a standard Merton-type portfolio problem.

Figure 4 presents the reduced value function, v , measured at the time of any reinvestment, i.e., $v(1)$, plotted against the asset's expected rate of return, μ , and standard deviation, σ . The default utility parameters are $\lambda = 2$, $\delta = 5\%$, $\beta = 0.3$, $\alpha_G = \alpha_L = 0.5$. For the μ and σ graphs, the other parameter is set to $\sigma = 30\%$ or $\mu = 9\%$, respectively. For comparison purposes, each value function is normalized to 1 at the values $\mu = 9\%$ and $\sigma = 30\%$.¹⁷

¹⁷ Standard utility functions are defined only up to a positive affine transformation. Realization utility has its level set so that a gain of zero gives a utility of zero, but scaling is still arbitrary.

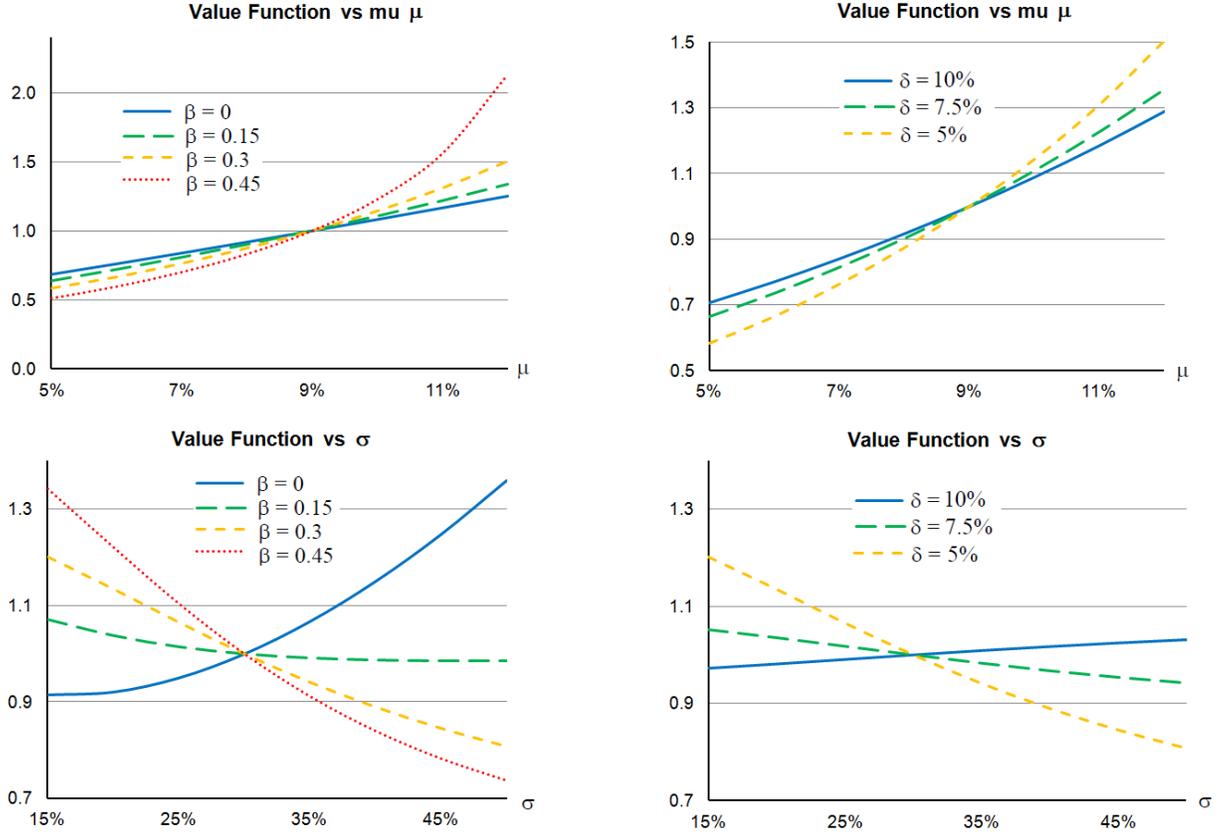


Figure 4: The Value Function. The initial optimized value function for scaled-TK utility plotted against μ and σ . The default parameters are $\mu = 9\%$, $\sigma = 30\%$, $k_s = k_p = 1\%$, $\beta = 0.3$, $\lambda = 2$, $\delta = 5\%$, $\alpha_G = \alpha_L = 0.5$. The investor fully recognizes the reinvestment cost in assessing his reinvestment gains, i.e., $\kappa = K \equiv (1-k_s)/(1+k_p)$.

Naturally the value function is increasing in μ . On average for higher μ , the next trade is more likely to be a gain and to occur sooner. The relation is steeper for a larger β because the continuation value from the reinvestment is larger. The relation is also steeper for smaller δ since the benefits from future gains are discounted less heavily.

Surprisingly, the value function is not always strictly decreasing in volatility as it is for a standard expected utility maximization; it can be increasing or U-shaped. Of course, CPT investors are risk-seeking with respect to losses, but that is not the reason for this effect. For example, the value function is increasing in volatility in the Barberis and Xiong (2012) model where burst utility is piece-wise linear and weakly concave. In our model there are conflicting effects.

Changing the three parameters, μ , δ , and σ^2 proportionally is identical to a change in the unit of time and leaves our model unaffected. So an increase in σ^2 can be interpreted as a proportional decrease in both μ and δ . Decreasing μ lowers the value function as just explained, but decreasing δ raises the value function since the future net positive bursts are discounted less heavily. As explained above, the smaller is β , the less important is the μ effect so for a small β the value function is less steeply decreasing or even increasing in volatility. Also the larger is δ

the more important is its effect. So for large δ , the value function is again less steeply decreasing or even increasing in volatility.

Figure 5 presents graphs of the optimal selling points for gains and losses for scaled-TK investors. The parameters left unchanged in each graph are set to the default values $\mu = 9\%$, $\sigma = 30\%$, $k_s = k_p = 1\%$, $\alpha_G = \alpha_L = 0.5$, $\beta = 0.3$, $\lambda = 2$, $\delta = 5\%$. The dotted blue lines show the optimal policies for an investor who ignores the reinvestment cost in assessing his gains, i.e. $\kappa = 1 - k_s$. The solid red lines show the optimal policies for an investor who does recognize these costs, i.e., $\kappa = (1 - k_s)/(1 + k_p)$. Several features are immediately evident.

For both types of investors, realized losses are typically much larger than realized gains so the basic strategy is to realize a few large losses and many small gains as we have already suggested intuitively. However, the no-sales region is wider for an investor who recognizes the reinvestment cost as reducing his gain. An investor who internalizes the costs more when assessing his well-being is obviously more reluctant to trade.

The upper sales point, Θ , is much less affected by parameter changes than is the lower sales point, θ , in most cases. In fact, Θ is largely unaffected by any of the variables except transaction costs and the scaling parameter, β . And for β , any effect occurs mostly near the transversality limit. As β approaches its limit of γ_1 , θ drops discontinuously to zero, and Θ approaches $\beta/[\kappa(\beta - \alpha_G)]$.¹⁸

On the other hand, the optimal loss-taking strategy does vary substantially as the utility parameters change. Increasing λ makes losses hurt more so θ is lowered in response to avoid some of them. Smaller α_L leads to higher marginal disutility for small losses. This induces the investor to wait longer to realize a loss as his risk-seeking behavior increases (α_L decreases from 1 to 0). Conversely for small α_G , the marginal utility of small gains is quite high making small losses “affordable” and desirable to set up future gains. So θ decreases with α_G . Similarly as β increases, the continuation utility, $(K\theta)^\beta v(1)$, falls. This increases the investor’s reluctance to experience a loss just to increase future gains.

The effect of the subjective discount rate, δ , on the optimal sales policy is unusual. A more impatient investor wants to realize gains sooner and defer losses longer. We see that Θ is decreasing in δ as expected; however, θ is not. All losses are taken voluntarily, and the desire to take gains early induces a derived willingness to realize losses in order to set up these future gains. This causes θ to also be increasing in δ at low discount rates; however, at higher discount rates the initial intuition dominates because future gains are discounted more so the benefit of resetting is less.

Increasing the transaction costs, k_s and k_p , naturally widens the no-sales region because the costs take part of each gain and increase every loss. As the costs go to zero, both Θ and θ approach 1, and the trading frequency increases without limit. Because marginal utility becomes

¹⁸ For $\beta > \gamma_1$, there is no well-defined optimal upper sale point. Any $\Theta > K^{-\beta/(\beta-\gamma_1)}$ provides infinite expected utility. See the Appendix for details on this and other transversality-type violations. When α_G is very close to 1, the upper sales point is sensitive to β and can be decreasing for low β .

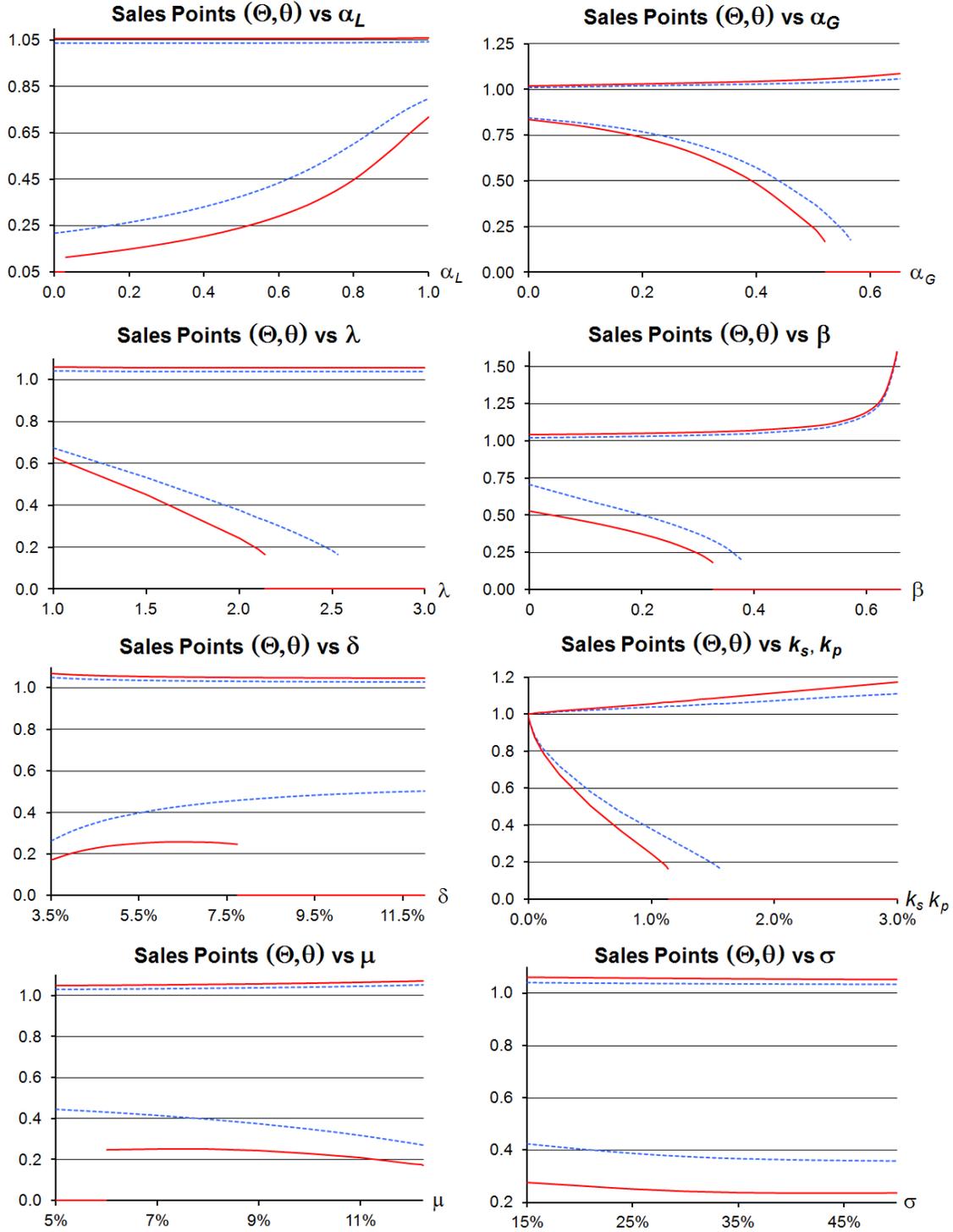


Figure 5: The Optimal Sales Policies for Scaled-TK Utility. The optimal sales points, Θ and θ , for scaled-TK utility plotted against various parameters. The default parameters in each graph are $\mu = 9\%$, $\sigma = 30\%$, $k_s = k_p = 1\%$, $\alpha_G = \alpha_L = 0.5$, $\beta = 0.3$, $\lambda = 2$, $\delta = 5\%$. The dotted blue lines show the optimal policies for an investor who ignores the reinvestment cost in assessing his gains, $\kappa = 1 - k_s$. The solid red lines show the optimal policies for an investor who recognizes the reinvestment cost in assessing his gains, $\kappa = K = (1 - k_s)/(1 + k_p)$.

infinite as the gain size goes to zero, the investor takes every opportunity to realize even the smallest of gains. Of course, there is unbounded marginal disutility for near-zero losses, but under the optimal strategy θ approaches 1 slower than does Θ , so there is a net increase in the value function.

There is a similar result for any utility function that is strictly concave for gains even if marginal utility is not infinite at zero. In the absence of transaction costs, gains are realized as soon as possible. This is always the optimal strategy, because the marginal utility is highest at 0 so any realized gain would always have greater utility if taken in smaller increments, and without transaction costs, there is no selling penalty to offset this. However, if the marginal utility is not infinite at zero, then the loss sales point need not be $\theta = 1$. In the presence of loss aversion ($\lambda > 1$), the marginal utility of small losses generally exceeds the marginal utility of small gains; this precludes an immediate loss realization in the absence of transaction cost.

Changes in μ have very little effect on size of optimal realized gains, Θ . The value function is strongly increasing in μ , but this is due to the reduction in the average time before a sale occurs rather than any change in policy. The lower sales point, θ , is affected more. For large or small μ , the option to reset the reference point by selling at a loss is less valuable than for intermediate values of μ . It is less valuable for large μ because the asset price grows quickly enough for gains to be realized without a painful reset. Conversely, with a very low μ , there is less value in resetting the reference level since future gains will be realized infrequently. So θ is highest for intermediate values of μ .

The standard deviation, σ , has only a tiny impact on the optimal gain point. Θ decreases as σ rises but by amount imperceptible in the graphs. Increasing σ also lowers the loss sales point, θ . The reason is, that for typical parameter values, higher volatility would increase the probability of loss realization, and the investors responds by lowering θ to postpone voluntary loss taking.

4. Model Calibration

In this section we calibrate our model using a representative investor to determine if it can explain observed trading patterns. The model described in the previous sections makes specific predictions about the magnitudes, frequencies, and relative proportion of both realized and unrealized gains and losses. In this context we explore an alternate utility specification which improves the model predictions. We also compare our model to a naïve prediction assuming random trading. The best calibration is achieved when we consider heterogeneous trading strategies.

If we consider a single set of utility and stock price parameters, our model predicts that the magnitudes of all realized gains and losses are $\Theta - 1$ and $\theta - 1$, respectively. The frequency of trading is determined by the time required for the investment value to rise to ΘR or fall to θR from the original reference level R . Paper gains expressed as a percentage of the reference level are distributed over the range $\theta - 1$ to $\Theta - 1$. The properties of these distributions are given in

Proposition 2. Its proof and that of all later propositions are provided in the Appendix.

Proposition 2: Properties of Realized and Unrealized Gains and Losses. If the asset value has a lognormal evolution, $dX/X = \mu dt + \sigma d\omega$, and the investor trades repeatedly according to a constant two-point policy, (θ, Θ) , then the probabilities that a given episode eventually ends with a gain or a loss are¹⁹

$$Q_G = \frac{\theta^{-\eta} - 1}{(\Theta/\theta)^\eta - 1} \quad Q_L = 1 - Q_G = \frac{1 - \Theta^{-\eta}}{1 - (\theta/\Theta)^\eta} \quad (13)$$

where $\eta \equiv 1 - 2\mu/\sigma^2$.

The fractions of the time the asset has an unrealized gain or loss are

$$\varphi_G = \frac{(1 - \theta^\eta)[\ell n \Theta - \frac{1}{\eta}(\Theta^\eta - 1)]}{(1 - \theta^\eta)\ell n \Theta + (\Theta^\eta - 1)\ell n \theta} \quad \varphi_L = 1 - \varphi_G. \quad (14)$$

The expected duration of each investment episode is

$$\mathbb{E}[\tau] = \frac{(\Theta^\eta - 1)\ell n \theta - (\theta^\eta - 1)\ell n \Theta}{(\Theta^\eta - \theta^\eta)(\mu - \frac{1}{2}\sigma^2)}, \quad (15)$$

and in a sequence of consecutive investments, the expected number of investment episodes per unit time is $1/\mathbb{E}[\tau]$. ■

The properties derived in this proposition and Proposition 4 below do not depend on the specific realization utility function assumed nor even on the maximization of utility at all. They obtain whenever a constant two sale point strategy is employed for assets with lognormal diffusions.

For comparison we derive the same predictions from a naïve model which assumes random trading for investors who may trade for liquidity purposes, based on information, or for other reasons. Describing the actions of these investors is outside the scope of this paper so we simply assume that these investors also trade stocks in separate episodes, however, each episode is terminated at random independently of the stock price evolution. The predictions of this model are given in Proposition 3.

Proposition 3: Characteristics of Investment Episodes for Random Trades. Assume that each asset's price evolves according to a lognormal diffusion and that each trading episode terminates with a sale that occurs according to a Poisson process which is independent of the evolution of the stock price and has intensity ρ . The trading episodes have the properties given below.

¹⁹ When $2\mu = \sigma^2$ (so $\eta = 0$), L'Hôpital's rule gives $Q_G = 1 - Q_L = \varphi_L = 1 - \varphi_G = -\ell n \theta / \ell n(\Theta / \theta)$ and $\mathbb{E}[\tau] = -\ell n(\theta)\ell n(\Theta) / 4\sigma^2$.

The duration of each episode has an exponential distribution with mean duration $\mathbb{E}[\tau] = 1/\rho$. The average realized gain and loss are²⁰

$$\bar{\Theta} = -\frac{\rho(1-\psi^-)}{(\rho-\mu)\psi^-} \quad \bar{\theta} = \frac{\rho(\psi^+ - 1)}{(\rho-\mu)\psi^+} \quad (16)$$

where $\psi^\pm \equiv \frac{-(\mu - \frac{1}{2}\sigma^2) \pm \sqrt{(\mu - \frac{1}{2}\sigma^2)^2 + 2\rho\sigma^2}}{\sigma^2}$.

The probability that a given episode will end with a sale at a gain and the probability that an unrealized investment is a paper gain are both

$$Q_G = \varphi_G = \frac{\psi^-}{\psi^- - \psi^+}. \quad (17)$$

Of course, the probability that a given episode will end in a loss and the probability of an unrealized loss are $Q_L = \varphi_L = 1 - Q_G = 1 - \varphi_G$. ■

Note that the trading points, $\bar{\Theta}$ and $\bar{\theta}$, given in (16) are averages. While threshold-realization-utility investors always trade at fixed ratios, the sales of a random trader can occur at any price.

One set of trading statistics that has garnered considerable attention were proposed by Odean (1998) to measure the disposition effect. These include the proportion of gains realized, *PGR*; the proportion of losses realized, *PLR*; and the Odean measure, \mathcal{O} . *PGR* is defined as a ratio. The numerator is the number of realized gains totaled over all days and all accounts. The denominator is the total over all days of the number of stock positions showing a gain (realized or not) in all accounts which realized either a gain or a loss on that day. *PLR* is similarly defined for losses. The Odean measure is the ratio PGR/PLR .

In a given sample, these statistics will be affected by many factors. These include the varying sales thresholds for the distinct assets held by different investors, the number of stocks held in each account, the correlation between the stocks' returns, and the sampling interval.²¹ Proposition 4 derives *PGR*, *PLR*, and the Odean measure for the special case of independent and identically distributed assets. The number of stocks held per account may vary across investors, but each individual stock is traded according to the same two-point or random strategy.

Proposition 4: Odean's Statistics with a Representative Investor. Assume that asset returns are independent and identically distributed with lognormal evolutions, $dX/X = \mu dt + \sigma d\omega$, and that all stocks are traded using the same strategy. Then as the number of sales increases, the

²⁰ For $\rho = \mu$, L'Hôpital's rule gives $\bar{\theta} = \mu / (\mu + \frac{1}{2}\sigma^2)$. The expected value of the upper sales, $\bar{\Theta}$, is finite only if $\rho > \mu$; $\bar{\theta}$ is always finite since its realizations are bounded between 0 and 1.

²¹ These factors are examined in Ingersoll and Jin (2012).

probability limits of PGR , PLR , and the Odean measure are

$$\begin{aligned} \text{plim } PGR &= \frac{Q_G}{Q_G + (\bar{n} + \sigma_n^2 / \bar{n} - 1)\varphi_G} & \text{plim } PLR &= \frac{Q_L}{Q_L + (\bar{n} + \sigma_n^2 / \bar{n} - 1)\varphi_L} \\ \text{plim } \mathcal{O} &= \frac{\text{plim } PGR}{\text{plim } PLR} = \frac{Q_G}{Q_L} \frac{Q_L + (\bar{n} + \sigma_n^2 / \bar{n} - 1)\varphi_L}{Q_G + (\bar{n} + \sigma_n^2 / \bar{n} - 1)\varphi_G} \end{aligned} \quad (18)$$

where \bar{n} and σ_n^2 are the average number and variance of the number of stocks held across accounts,²² and Q_G , Q_L , φ_G , and φ_L are the probabilities defined in (13) and (14) if the investors are realization-utility investors or in (17) if the investors are for random Poisson traders. ■

The statistics of Proposition 4 are the same as those that would be produced by a single representative investor holding $\bar{n} + \sigma_n^2 / \bar{n}$ rather than \bar{n} stocks. The representative investor's holding is biased high relative to the average because those investors holding more stocks are represented more often in the data.

With the statistics derived in Propositions 2 through 4, we can assess how our realization utility model and the naïve model fit the trading patterns observed for individual traders. For comparison we incorporate data from Odean (1998) with the statistics generated by this model into Table 1. The first row of the table presents Odean's data extracted from his Tables 1 and 3 including the header text.

To determine if any calibration is feasible, the data in the second row uses just Odean's average sales price ratios as estimates for Θ and θ . The remaining values are determined using Propositions 2 and 4. This fit is not optimized; we have simply chosen an asset comparable to a typical share of stock with $\mu = 9\%$ and $\sigma = 30\%$. The fit for Q_G and φ_G , and therefore the corresponding loss statistics, do seem reasonable allowing for sampling error and heterogeneity of assets and investors in the actual sample.²³ For a similar data set, Barber and Odean (2000) report that the average number of stocks per account is 4.3. The best quadratic fit for PGR and PLR is about $\bar{n} + \sigma_n^2 / \bar{n} = 8.0$ implying $\sigma_n = 3.99$ which seems a realistic estimate of the variation in account sizes and is consistent with the estimate from Goetzmann and Kumar (2008).²⁴

²² If each account holds a fixed number of assets over time, then \bar{n} and σ_n^2 are the average and variance of account sizes. Under mild regularity conditions, the proposition remains valid for accounts whose sizes vary over time with \bar{n} and σ_n^2 the average and variance of the number of shares held per account across both accounts and time.

²³ The estimated average holding time of 312 trading days is based on a 250 trading-day year and assumes Odean's reported average holding time of 15 months is exact. The actual value could be 302 to 323 days, but is substantially larger than our fitted Odean value of 174 days. This difference could be due to the representative stock chosen. All of the statistics in the last six columns are invariant to time scaling except for $\mathbb{E}[\tau]$; the others depend only on the ratio μ/σ^2 . In our analysis below it is only the *relative* holding times for different accounts that matters not the level.

²⁴ Goetzmann and Kumar (2008), using the same data set as in Barber and Odean (2000), provide more details about portfolio sizes. They give the percentages of accounts holding 1, 2, 3, 4, 5, 6-10, 11-15, >15 stocks. Their Table 1 confirms that the standard deviation is approximately 4.

Table 1: Summary Statistics for Reference-Dependent Realization Utility Model with Scaled Tversky-Kahneman Utility

The table reports: $\Theta - 1$, $\theta - 1$: percentages above and below the reference level for realized gains and losses, Q_G : fraction of episodes that end in realized gains, $\mathbb{E}[\tau]$: average holding period in trading days (250 per year), φ_G : fraction of stocks with unrealized paper gains, PGR , PLR : percentage of gains and losses realized, and $\mathcal{O} \equiv PGR/PLR$: Odean's measure. Asset parameters are $\mu = 9\%$ and $\sigma = 30\%$. The accounts' sizes are fixed with $\bar{n} + \sigma_n^2/\bar{n} = 8.0$. Utility parameters are $\lambda = 2$ and $\delta = 5\%$ (except $\delta = 8\%$ for $\alpha_G = 0.88$ and $\delta = 10\%$ for $\alpha_G = 1$ to avoid a transversality violation. Transaction costs are $k_s = k_p = 1\%$, and the investor accounts for both costs in his subjective view of his realized gain. Odean's data is taken from Tables 1 and 3 of his 1998 paper including the header description. The "Fit to Odean's Θ , θ " row uses Odean's estimates of Θ and θ and computes the other values using Propositions 2 and 4. Each "Poisson Model" row chooses ρ to match one of Odean's estimates of Θ , θ , Q_G , or $\mathbb{E}[\tau]$ and computes the other values using Propositions 3 and 4; the observed φ_G cannot be matched as the Poisson model can only give values less than 50% when $\mu < \sigma^2/2$.

	$\Theta - 1$	$\theta - 1$	Q_G	φ_G	$\mathbb{E}[\tau]$	PGR	PLR	\mathcal{O}
Odean's data	27.7%	-22.8%	53.8%	41.9%	312	14.8%	9.8%	1.51
Fit to Odean's Θ , θ	27.7%	-22.8%	57.7%	50.7%	174	14.0%	10.9%	1.28

Random Trading (Poisson) Model

$\rho = 0.36$	72.2%	-22.8%	58.7%	58.7%	688	12.5%	12.5%	1
$\rho = 0.80$	36.4%	-17.4%	55.9%	55.9%	312	12.5%	12.5%	1
$\rho = 1.16$	27.7%	-15.2%	54.9%	54.9%	215	12.5%	12.5%	1
$\rho = 1.94$	19.7%	-12.4%	53.8%	53.8%	129	12.5%	12.5%	1

Realization Model with Scaled-TK Utility

$\alpha_G = 1$	$\beta = 0,1$	95.3%	never	100%	27.1%	3717	34.5%	0	∞
$\alpha_L = 1$	$\beta = 0.53$	45.6%	never	100%	16.6%	2087	46.2%	0	∞
$\alpha_G = 0.88$	$\beta = 0$	17.6%	never	100%	7.7%	901	65.0%	0	∞
$\alpha_L = 0.88$	$\beta = 0.88$	96.2%	never	100%	27.3%	3743	34.4%	0	∞
$\alpha_G = 0.5$	$\beta = 0$	3.9%	-13.5%	80.6%	21.5%	15	34.9%	3.4%	10.22
$\alpha_L = 0.88$	$\beta = 0.3$	5.8%	-45.3%	93.8%	9.5%	85	58.6%	1.0%	60.64
$\alpha_G = 0.5$	$\beta = 0$	3.8%	-6.3%	64.9%	36.7%	7	20.2%	7.3%	2.74
$\alpha_L = 1.0$	$\beta = 0.3$	5.9%	-28.2%	87.6%	15.6%	50	44.5%	2.1%	21.64
$\alpha_G = 0.5$	$\beta = 0$	4.0%	-47.3%	95.9%	6.5%	63	67.8%	0.6%	107.90
$\alpha_L = 0.5$	$\beta = 0.3$	5.7%	-75.8%	98.3%	4.9%	169	74.3%	0.3%	293.06

The next four rows of the table illustrate the fit of a trading model based on random Poisson trades to Odean's data. As ρ increases, both average sales points, $\bar{\Theta}$ and $\bar{\theta}$, approach 1. Under random trading, Q_G and φ_G must be equal, and both fall from 100% to 50% as ρ increases from 0 to ∞ .²⁵ While the individual statistics can be matched, they cannot be fit simultaneously. In addition, PGR and PLR must both equal $(\bar{n} + \sigma_n^2/\bar{n})^{-1}$ with random trading so the Odean measure must always be 1.

The final ten rows of the table attempt to fit Odean's data to a specific realization utility function. Using $\alpha_G = \alpha_L = 0.88$ and $\lambda = 2.25$, as proposed by Tversky and Kahneman (1992), and $\delta = 8\%$, the upper sales point of our model varies from 17.6% to 96.2% above the reference level as the scaling parameter, β , ranges over its permitted values 0 to 0.88. This does include Odean's estimate of 27.7%, but the fit is far from satisfactory as no losses are voluntarily realized for these parameters.

For piecewise linear utility, $\alpha_G = \alpha_L = 1$ (and $\delta = 10\%$ to avoid transversality violation), the table confirms that voluntary loss taking is again never optimal as was shown in Proposition 1. Gains are realized when after the stock price has risen by 95% for $\beta = 0$ or 1 or by a lesser amount for any β between those values. The smallest gain-realization point is 45.6% when $\beta = 0.53$. It is apparent from the table that realization utility model cannot fit Odean's data with α values this high.

Tversky and Kahneman estimates are from experimental settings with small gamble sizes. For the much larger size of investment that a typical investor makes in financial markets we expect more risk aversion.²⁶ Therefore, in Table 1 we also use $\alpha_G = 0.5$ which is the estimate of Wu and Gonzalez (1996).²⁷ This also permits lowering the rate of time preference to a more reasonable 5%. Since Wu and Gonzalez only estimated α_G , we use α_L in the range 0.5 to 1.0

From Table 1 it is apparent that the basic model can generate a wide variety of optimal sales points; however, for any parameter values that permit voluntary sales at the size of losses observed in Odean's data, the upper sales point, Θ , is much too low. As a direct result, sales at gains vastly outnumber sales at losses ($Q_G > Q_L$) and PGR is too large while PLR is too small.

The fundamental problem with scaled-TK is that marginal utility is very high for small gains, indeed marginal utility is infinite at 0 for any $\alpha_G < 1$. This substantially increases the benefit of realizing gains in numerous small increments making the optimal threshold, Θ , quite small.

²⁵ If $2\mu < \sigma^2$, then both Q_G and φ_G rise from 0 to 50% as ρ increases. In this case Odean's value of $Q_G = 53.8\%$ cannot be matched. Conversely, for the parameters used here, φ_G cannot be matched to Odean's value of 41.9%.

²⁶ Barber and Odean (2000) report that the average household in their sample holds 4.3 stocks worth \$47,334, so the averaged dollar amount invested per stock a little more than \$11,000. In Table 6 of Tversky and Kahneman, the largest gamble about which subjects were questioned was \$401 which represents a modest 3.6% gain on the average stock position.

²⁷ Wu and Gonzalez (1996) only estimate α_G . Their estimation depends on the form for the probability weighting function. When using that proposed in Tversky and Kahneman (1992), they estimate $\alpha_G = 0.5$; using the form proposed in Prelec (1998), they estimate of $\alpha_G = 0.48$.

To avoid this problem, we consider a modified-TK utility function

$$U_{\text{mKT}}(G, R) = \begin{cases} R^\beta [(1 + G/R)^{\alpha_G} - 1] / \alpha_G & G \geq 0 \\ -\lambda R^\beta [1 - (1 + G/R)^{\alpha_L}] / \alpha_L & -R \leq G < 0. \end{cases} \quad (19)$$

For $\alpha_G < 1 < \alpha_L$, utility is S-shaped; otherwise the risk parameters, α_G and α_L , are unbounded below and above, respectively, allowing more flexibility.²⁸ Marginal utility is bounded at $G = 0$. It reaches the values $\lambda R^{\beta-1}$ and $R^{\beta-1}$ just below and above zero with a discontinuous change inducing a true kink in the utility function. In addition, for scaled-TK utility the utility burst for losing the entire reference level is $-\lambda R^\beta$ which does not depend on the risk seeking parameter α_L while for modified-TK utility, the utility burst for losing R is $-\lambda R^\beta / \alpha_L$ which increases with α_L .

Table 2 provides additional calibrated results using the modified-TK utility specification. The results presented there are not as extreme as those in Table 1 using scaled-TK utility. In particular, the lower marginal utility at small gains has served its purpose of raising the optimal gains threshold, Θ . The estimates in the final row match the data quite well. As should be obvious from the parameters, this is not an optimized or best fit; rather round numbers are used for α_G , α_L , and β which provide a good fit.

It might well be argued that $\alpha_L = 30$ implies an implausibly high risk-seeking behavior, so the model is doubtful despite fitting the data. Further parameter adjustments cannot do much to improve the fit. For a given threshold-sales policy, Θ - θ , the remaining values in Tables 1 and 2, except the average holding time, are completely determined by the ratio μ/σ^2 . Therefore, adjusting the utility parameters further cannot better the fit, nor can altering μ or σ improve the fit for PGR and PLR without degrading that for Q_G and φ_G . However, the calibration can be improved by introducing additional heterogeneities beyond a difference in the number of stocks held because this model is not one in which a single average investor can serve as a stand-in for the group.

If investors trade some of their stocks differently (*heterogeneous holdings*) or different investors have different sales policies (*heterogeneous investors*), there are further aggregation effects on the various measures. In particular, the closer are Θ and θ to one, the shorter will be the average length of each investment episode. Stocks that are traded more frequently will disproportionately affect the statistics because their characteristics and those of other stocks held in the same accounts will be over-represented. The effects of aggregation are described in Propositions 5 and 6. Proposition 5 gives the statistics when different investors follow distinct trading strategies. Proposition 6 gives the statistics when the trading strategies differ for stocks within the same account. These heterogeneities have distinct effects.

²⁸ As usual $\alpha = 0$ corresponds to a logarithmic form, $R^\beta \ln(1+G/R)$. Modified-TK utility can also be adapted to study strictly risk-averse incremental utility by setting $\alpha_L < 1$. Utility is increasing and for $\lambda \geq 1$ strictly concave. If $\alpha_L = \alpha_G$ and $\lambda = 1$, this is incremental power utility; otherwise there is a discontinuous change in risk aversion (if $\alpha_G \neq \alpha_L$) or in marginal utility (if $\lambda \neq 1$) at 0.

Table 2: Summary Statistics for Reference-Dependent Realization Utility Model with Modified Tversky-Kahneman Utility

The table reports: $\Theta - 1$, $\theta - 1$: percentages above and below the reference level for realized gains and losses, Q_G : fraction of episodes that end in realized gains, $\mathbb{E}[\tau]$: average holding period in trading days, φ_G : fraction of stocks with unrealized paper gains, PGR , PLR : percentage of gains and losses realized, and $\mathcal{O} \equiv PGR/PLR$: Odean's measure. Asset parameters are $\mu = 9\%$ and $\sigma = 30\%$. The account's sizes are fixed with $\bar{n} + \sigma_n^2/\bar{n} = 8.0$. Utility parameter are $\lambda = 2$ and $\delta = 5\%$. Transaction costs are $k_s = k_p = 1\%$, and the investor accounts for both costs in his subjective view of his realized gain. Odean's data is taken from Tables 1 and 3 of his 1998 paper including the header description. The "Fit to Odean's Θ , θ " row uses Odean's estimates of Θ and θ and computes the other values using Propositions 2 and 4.

	$\Theta - 1$	$\theta - 1$	Q_G	φ_G	$\mathbb{E}[\tau]$	PGR	PLR	\mathcal{O}
Odean's data	27.7%	-22.8%	53.8%	41.9%	312	14.8%	9.8%	1.51
Fit to Odean's Θ , θ	27.7%	-22.8%	57.7%	50.7%	174	14.0%	10.9%	1.28

Realization Model with Modified-TK Utility									
$\alpha_G = 0.5$	$\beta = 0$	60.4%	-90.7%	96.3%	25.2%	2037	35.3%	0.7%	50.25
		$\alpha_L = 2$	$\beta = 0.3$	49.2%	Never	100.0%	17.6%	2220	44.8%
$\alpha_G = 0.5$	$\beta = 0$	44.6%	-64.1%	85.3%	31.5%	909	27.9%	3.0%	9.36
		$\alpha_L = 4$	$\beta = 0.3$	47.4%	-73.6%	89.7%	28.3%	1169	31.1%
$\alpha_G = 0.5$	$\beta = 0$	27.5%	-42.7%	77.6%	33.3%	351	25.0%	4.6%	5.44
		$\alpha_L = 8$	$\beta = 0.3$	38.3%	-48.7%	77.4%	36.5%	556	23.2%
$\alpha_G = 0.5$	$\beta = 0$	13.5%	-17.5%	64.1%	41.0%	67	18.3%	8.0%	2.29
		$\alpha_L = 30$	$\beta = 0.3$	26.7%	-24.3%	60.5%	48.0%	181	15.2%

Proposition 5: Realization Utility Statistics with Heterogeneous Investors. Assume that asset returns are independent and identically distributed²⁹ and that investors differ in their trading strategies or number of stocks they hold. Type i investors constitute the fraction π_i of the sample, hold n_i stocks, and follow a Θ_i - θ_i threshold strategy or a ρ_i random strategy. As the number of the observed trades increases, the probability limits of the various aggregate statistics are the weighted averages

$$\text{plim } \bar{\Xi} = \sum_i w_i \Xi_i \quad \text{where} \quad w_i \equiv \frac{\pi_i n_i / \mathbb{E}[\tau_i]}{\sum_i \pi_i n_i / \mathbb{E}[\tau_i]} \quad (20)$$

and Ξ is any of the statistics Θ , θ , Q_G , Q_L , or $\mathbb{E}[\tau]$. The probability limits of the fraction of

²⁹ The assets' means and variances can differ across the types of investors, provided they are identical within types. The effects of any asset differences are completely incorporated into Q_G^i , Q_L^i , φ_G^i , φ_L^i , and $\mathbb{E}[\tau_i]$.

unrealized paper gains or losses are

$$\text{plim } \bar{\varphi}_G = 1 - \text{plim } \bar{\varphi}_L = \frac{\sum_i \pi_i n_i (n_i - 1) \varphi_G^i / \mathbb{E}[\tau_i]}{\sum_i \pi_i n_i (n_i - 1) / \mathbb{E}[\tau_i]}. \quad (21)$$

The probability limits of PGR , PLR , and the Odean measure are

$$\text{plim } PGR = \frac{\sum_i \pi_i n_i Q_G^i / \mathbb{E}[\tau_i]}{\sum_i \pi_i n_i [Q_G^i + (n_i - 1) \varphi_G^i] / \mathbb{E}[\tau_i]} \quad \text{plim } PLR = \frac{\sum_i \pi_i n_i Q_L^i / \mathbb{E}[\tau_i]}{\sum_i \pi_i n_i [Q_L^i + (n_i - 1) \varphi_L^i] / \mathbb{E}[\tau_i]}. \quad (22)$$

As before, the probability limit for the Odean measure is the ratio of PGR to PLR . ■

Proposition 6: Realization Utility Statistics with Heterogeneous Holdings. Assume that a representative investor trades N stocks whose returns are independently distributed. These stocks are grouped into categories. Within group i , there are n_i stocks with identical means, variances and trading strategies. The latter are either a Θ_i - θ_i threshold strategy or a ρ_i random strategy. As the number of the observed trades increases, the probability limits of the various aggregate statistics are the weighted averages

$$\text{plim } \bar{\Xi} = \sum_i w_i \Xi_i \quad \text{where} \quad w_i \equiv \frac{n_i / \mathbb{E}[\tau_i]}{\sum_i n_i / \mathbb{E}[\tau_i]} \quad (23)$$

and Ξ is any of the statistics Θ , θ , Q_G , Q_L , or $\mathbb{E}[\tau]$. The probability limits of the fraction of unrealized paper gains or losses are

$$\text{plim } \bar{\varphi}_G = 1 - \text{plim } \bar{\varphi}_L = \frac{\sum_i n_i (\sum_j n_j \varphi_G^j - \varphi_G^i) / \mathbb{E}[\tau_i]}{(N - 1) \sum_i n_i / \mathbb{E}[\tau_i]}. \quad (24)$$

The probability limits of PGR , PLR , and the Odean measure are

$$\text{plim } PGR = \frac{\sum_i n_i Q_G^i / \mathbb{E}[\tau_i]}{\sum_i n_i [Q_G^i + \sum_j n_j \varphi_G^j - \varphi_G^i] / \mathbb{E}[\tau_i]} \quad \text{plim } PLR = \frac{\sum_i n_i Q_L^i / \mathbb{E}[\tau_i]}{\sum_i n_i [Q_L^i + \sum_j n_j \varphi_L^j - \varphi_L^i] / \mathbb{E}[\tau_i]} \quad (25)$$

with $\text{plim } \mathcal{O} = \text{plim } PGR / \text{plim } PLR$ as always. ■

Table 3 summarizes the calibrated results for heterogeneous investors and heterogeneous holdings based on the statistics derived in Proposition 5 and 6. For comparison purpose, we pick the utility parameters used in Table 2, excluding the high risk tolerance case which is no longer needed. All of the averages in this table assume there is an equal mix of two types. For “heterogeneous investors,” one half of the investors optimize realization utility and the other half trade randomly. For “heterogeneous holdings,” each investor trades one-half of his stocks by

Table 3: Summary Statistics for Mixed Reference-Dependent Realization Utility and Random Trading

The table reports: $\Theta-1$, $\theta-1$: percentages above and below the reference level for realized gains and losses, Q_G : fraction of episodes that end in realized gains, $\mathbb{E}[\tau]$: average holding period in trading days, ϕ_G : fraction of stocks with unrealized paper gains, PGR , PLR : percentage of gains and losses realized, and $\mathcal{O} \equiv PGR/PLR$: Odean's measure. Asset parameters are $\mu = 9\%$ and $\sigma = 30\%$. The account's sizes are fixed with $\bar{n} + \sigma_n^2/\bar{n} = 8.0$. In each row under Heterogeneous Investors, one half of the investors are realization-utility traders with utility parameters $\lambda = 2$, $\delta = 5\%$, and others as stated. The other half of the investors trade randomly with the stated Poisson intensity. In each row under Heterogeneous Holdings, each investor trades half his stocks to optimize realization utility and the other half randomly.

			$\bar{\Theta}-1$	$\bar{\theta}-1$	\bar{Q}_G	$\overline{\mathbb{E}[\tau]}$	Heterogeneous Investors				Heterogeneous Holdings			
							$\bar{\phi}_G$	PGR	PLR	\mathcal{O}	$\bar{\phi}_G$	PGR	PLR	\mathcal{O}
Odean's data			27.7%	-22.8%	53.8%	312	41.9%	14.8%	9.8%	1.51	41.9%	14.8%	9.8%	1.51
Fit to Odean's Θ, θ			27.7%	-22.8%	57.7%	174	50.7%	14.0%	10.9%	1.28	50.7%	14.0%	10.9%	1.28
$\alpha_G = 0.5$ $\alpha_L = 2$	$\beta=0$	$\rho=1.5$	26.1%	-19.5%	52.1%	308	52.1%	13.6%	11.3%	1.21	42.5%	17.8%	8.9%	1.99
		$\rho=1$	34.1%	-24.2%	52.0%	445	52.0%	14.1%	10.7%	1.32	40.2%	18.1%	8.6%	2.12
	$\beta=0.3$	$\rho=1.5$	25.1%	-19.7%	51.7%	310	51.7%	13.7%	11.2%	1.23	42.5%	19.6%	8.4%	2.34
		$\rho=1$	32.7%	-24.5%	51.5%	449	51.5%	14.2%	10.6%	1.35	40.2%	19.9%	8.0%	2.48
$\alpha_G = 0.5$ $\alpha_L = 4$	$\beta=0$	$\rho=1.5$	26.6%	-21.5%	50.8%	282	50.8%	14.3%	10.6%	1.34	40.9%	16.8%	9.1%	1.84
		$\rho=1$	33.8%	-26.4%	50.1%	392	50.1%	15.0%	9.9%	1.51	38.3%	17.2%	8.7%	1.99
	$\beta=0.3$	$\rho=1.5$	29.4%	-23.3%	50.9%	292	50.9%	14.4%	10.5%	1.38	41.3%	17.4%	8.9%	1.94
		$\rho=1$	33.8%	-26.2%	50.5%	412	50.5%	14.8%	10.0%	1.47	38.7%	17.8%	8.5%	2.09
$\alpha_G = 0.5$ $\alpha_L = 8$	$\beta=0$	$\rho=1.5$	24.6%	-23.0%	47.6%	226	47.6%	15.7%	9.4%	1.66	38.2%	16.9%	8.8%	1.93
		$\rho=1$	29.4%	-27.1%	46.1%	292	46.1%	16.7%	8.6%	1.94	35.5%	17.3%	8.3%	2.09
	$\beta=0.3$	$\rho=1.5$	26.7%	-21.8%	50.2%	256	50.2%	14.5%	10.4%	1.40	40.4%	16.0%	9.4%	1.69
		$\rho=1$	33.2%	-26.1%	49.5%	345	49.5%	15.2%	9.7%	1.57	37.9%	16.4%	9.0%	1.81

optimizing his realization utility and trades the other half randomly. The statistics, Θ , θ , Q_G , and $\mathbb{E}[\tau]$, depend only on the trading characteristics of all the stocks in aggregate and not on the specific holdings, so they are the same for both the heterogeneous investors and heterogeneous holdings cases. The other statistics differ in the two cases.

Following Odean, φ_G (φ_L) is the average number of stocks trading at a paper gain (loss) when some other stock in the same account is sold. Stocks with a lower $\mathbb{E}[\tau]$ trade more frequently and therefore increase the impact of all other stocks *held in the same account* on φ_G and φ_L and the other statistics that depend on these. For example from the seventh line in Table 2, investors with $\alpha_L = 8$, $\alpha_G = 0.5$, $\beta = 0$, have an average holding period of 351 trading days with $\varphi_G = 33.3\%$; $\rho = 1$ random trades have a smaller average holding period of 250 days and a higher $\varphi_G = 50.1\%$. Therefore, with an equal mixture of these investors, the average $\bar{\varphi}_G = 46.1\%$ is closer to that of the random traders since they trade more often. Conversely with heterogeneous holdings, the opposite is true. Since random trades occur more often, the threshold-traded stocks are the ones observed more often in determining the paper gains and losses. As a consequence PGR and \mathcal{O} are larger and PLR is smaller for heterogeneous holdings than for heterogeneous investors.

The reported averages are in fairly close agreement with Odean's empirical results for each level of risk tolerance with no need to resort to the very high risk tolerance required in Table 2. The average time between trades has also increased to better match Odean's value. This table simply highlights the possibilities. Using mixtures of other than 50-50 or including more types of investors would permit further tweaking of the fit. Combining Propositions 5 and 6 to reflect a more general case in which investors have a different number of stocks and trade stocks differently is not discussed here because of its minimal benefit for further improving the fit.

In summary, we have shown that our realization utility model is consistent with and can shed light on several dimensions of observed trading data. We have made no attempt to match the empirical patterns exactly, nor have we considered all dimensions. For instance, in his 1998 paper, Odean reported the average size of paper gains and losses measured when other stocks in the account were traded. The average paper gain and loss were 46.6% and -39.3%, which are larger in magnitude than the realized gains and losses of 27.7% and -22.8%. Our model with a single representative investor following an identical Θ - θ strategy on all the stocks cannot possibly generate this pattern as the realizations occur at the extreme points of the paper gains and losses distribution. Although it is possible to generate this pattern by considering an economy with heterogeneous investors, one other explanation is that investors update their reference levels based on the recent price history of the asset.³⁰ Once the reference level changes from the initial purchase price, the subjectively measured gain or loss realized by an investor differs from the gain or loss as measured objectively by an outside econometrician. This topic is examined in detail in Ingersoll and Jin (2012).

³⁰ See section IV of Odean (1999) for an informal discussion on this.

5. Voluntary Loss Realization

One of the main objectives of this paper is to create an intertemporal realization utility model with voluntary loss realization. In earlier sections we introduced scaled- and modified-TK utility functions, and our numerical analysis showed that either can generate voluntary loss taking. In this section, we pose a more general question: what are the necessary characteristics for *any* utility function to generate voluntary sales at losses in the intertemporal realization utility framework posed in (3)? We are not ruling out other possible preference or belief-based explanations that contribute to voluntary loss taking such as changes in information or portfolio rebalancing. Nevertheless, our general analysis may shed light on theoretical and experimental work on realization utility and the disposition effect.

Both versions of TK utility have two properties that separately seem important, an S-shape and scaling. To illustrate consider an investor with the scale-free utility function, $U(G) = \text{sgn}(G) \cdot |G|^{1/2}$ that does not explicitly depend on the reference level. This investor might be willing to realize a loss of 2 since, if the price subsequently recovers, he can take two gains of 1 and have positive total utility. Not taking the loss of 2 prevents realizing the recovery as a gain. With scale-free concave utility, the benefit of a recovery gain can never offset the disutility of the loss because marginal utility is decreasing so the disutility of the loss must be larger than the utility of subsequent gains no matter how it is divided. Now consider a simple scaled utility function with no convexity for losses, $U(G, R) = G/R$. A loss from 4 to 2 has a disutility of $-2/4$ while a recovery from 2 back to 4 provides a utility of $2/2$. Now the decreasing scaling may make taking the loss worthwhile in order to realize a later gain. Conversely, with increasing scaling, loss taking could never be optimal. This analysis is made precise in Proposition 7 below which describes the condition under which losses will never be realized voluntarily.

Proposition 7: Assume that an investor maximizes expected realization utility as in (3) with a burst utility function of $U(G, R)$ and that the reference level is updated to the investment value after a sale but remains constant between sales. The following four conditions

$$(i) \delta > 0 \quad (ii) \frac{\partial U}{\partial G} > 0 \quad (iii) \frac{\partial^2 U}{\partial G^2} \leq 0 \quad (iv) \frac{\partial^2 U}{\partial G \partial R} \geq 0 \quad (26)$$

are jointly sufficient to preclude the voluntary realization of any losses even in the absence of transaction costs.³¹ ■

A proof of this theorem is given in the Appendix. The intuition for the result is that the disutility of a loss cannot be offset by the utility of any later gains that recoup this loss because those utility contributions come at a later time, have a lower reference level, and smaller marginal utility. Each of these aspects makes the gain utility smaller by properties (i), (iv), and (iii), respectively.

³¹As seen in Figure 5 in section 3, transaction costs widen the no-sales region and therefore make voluntary loss taking less likely to occur. With transaction costs we would require stronger violations of one of the four conditions to make voluntary loss taking optimal.

The first three conditions in (26) are standard utility properties. Time preference is positive, and marginal utility is positive and decreasing. Taking the first two as inviolable, voluntary losses are then possible only if utility is not everywhere concave or marginal utility is not everywhere increasing in the reference level. We have already discussed the option-like effect that S-shaped utility has on loss taking from the resetting of the reference level. A violation of condition (iv), that is, marginal utility is decreasing in the reference level over some range, can have the same effect. A loss measured from a high R can have a smaller negative impact on utility than the same size gain from a lower R . This might induce the investor to realize a loss so he is in position to take a later gain when the asset recovers in value.

S-shaped utility functions are commonly used in behavioral models, and we have already argued that $\partial|U(G, R)|/\partial R \leq 0$ is a likely description of realization utility which, if true, means (iv) cannot hold.³² Therefore, both S-shaped utility and decreasing scaling of utility with respect to R both may contribute to the optimal voluntary realization of losses. In theory a violation of (iii) or (iv) alone is sufficient to make voluntary loss taking optimal. However, our numerical analysis indicates that both are necessary for reasonable model calibration.

6. Further Model Predictions and Applications

As discussed above, our paper makes several direct and specific predictions about trading activity.³³ The two-point, Θ - θ , trading strategy is quite specific about the volume of gains and losses, the holding periods, and the sizes of both realized and paper gains and losses. These results can help to explain the difference between the trading volume in rising and falling markets as well as the effect of historical highs on the propensity to sell. Together, risk-seeking behavior and the trading strategy might also explain the heavy trading of highly valued assets since the optimal strategies are related to the assets' means and variances. Furthermore, if an investor's subjective reference level is not constant but is updated based on recent stock prices, then the predicted trading patterns become path-dependent.

Models like this one may also rationalize hazard-rate types of models for investor behavior. Ben-David and Hirshleifer (2012) recently documented a V-shape empirical pattern between the probability of selling a stock and the unrealized return since purchase for fixed holding periods.³⁴ Using the statistics developed in Proposition 5 for heterogeneous investors, our model can match this pattern. Many of these considerations are examined further in Ingersoll and Jin (2012).

³² Letting subscripts denote partial derivatives, $U_2(G, R) = U_2(0, R) + \int_0^G U_{12}(g, R)dg$. Since $U(0, R) = 0$ for all R , $U_2(0, R) = 0$, so if U_{12} has a uniform positive sign, $\text{sgn}(U_2) = \text{sgn}(U_{12})$ for $G > 0$. Similar reasoning shows that for $G < 0$, $\text{sgn}(U_2) = -\text{sgn}(U_{12})$. Therefore property (iv) guarantee that $\partial|U|/\partial R \geq 0$.

³³ A more detailed discussion on some of these applications can be found in Barberis and Xiong (2012).

³⁴ Ben-David and Hirshleifer (2012) implicitly assume that investors trading is random with hazard rate (intensity ρ) that is a function of holding period and the size of the unrealized gain. In contrast in our model with heterogeneous investors, the representative investor is not an average investor, and the hazard rate measures an aggregation effect.

Tax-trading behavior is another obvious topic addressed by this model. In fact, capital gains are a near perfect fit with realization utility since they are typically due only upon sale of an asset. Standard reasoning indicates that an investor should realize losses and defer gains to minimize the tax burden. This is the opposite of the disposition effect, with taxes loss taking should predominate over gain taking. However, this reasoning again fails to recognize the importance of the reinvestment effect. To illustrate, consider an investor with modified-TK utility with $\alpha_G = 0.5$, $\alpha_L = 4$, $\beta = 0.3$, $\lambda = 2$, $\delta = 5\%$ investing in an asset with $\mu = 9\%$ and $\sigma = 30\%$ and paying 1% transaction costs. In the absence of a capital gains tax, the investor would sell at $\Theta = 1.460$ or $\theta = 0.261$.³⁵ With a 15% capital gains tax, he sells at $\Theta = 1.549$ or $\theta = 0.248$. The capital gains tax does cause the investor to postpone the realization of gains because they now provide a smaller utility burst. However, loss taking also is postponed precisely because gains are less valuable so it no longer is quite so advantageous to realize a painful loss to reset the reference point. Of course, loss taking is affected less than gain taking because of different tax treatment.

The model also makes other indirect predictions. Two such predictions are the flattening of the capital market line and the pricing of idiosyncratic risk. There is no equilibrium model in our paper so precise predictions are not possible, but the model does indicate directional effects.

The μ - σ indifference curves for a realization-utility investor are flatter than the observed capital market line and, in some cases, are actually decreasing with the investor preferring more variance to less. But our model does not address to what these indifference curves should be applied. If an investor holds only a diversified mutual fund, the model indicates he should display less risk aversion than typical in the selection of a fund. But it also seems plausible that investors might save the bulk of their wealth in diversified portfolios recognizing that those are the best vehicle for long-term saving and still actively trade other stocks because they enjoy doing so. Only this latter investment activity might be governed by realization utility of the type we have modeled. This might explain why we see only a small number of stocks in typical trading accounts.

In this case there will be an excess demand for high-variance stocks. This means that high-beta and high residual risk stocks should have smaller expected returns than predicted by equilibrium models like the CAPM which ignore this behavior. Ang, Hodrick, Xing, and Zhang (2006) documented this. They tested alphas for stocks with different total and residual variation and found just this result. The difference between the alphas of the highest and lowest volatility stocks was -1.35% ($t = -4.62$) for the CAPM and -1.19% ($t = -5.92$) for Fama-French three-factor model. The difference in alphas between the highest and lowest residual volatility stocks was -1.38% ($t = -4.56$) for the CAPM and -1.31 ($t = -7.00$) for FF-3.

Precise predictions are, unfortunately, not possible, because we have ignored modeling diversification directly. In a standard portfolio model, risk-averse investors optimally hold many assets to provide diversification. Our model assumes narrow framing with the utility from each

³⁵ For consistency in the comparison, we assumed that the investor views his subjective gain as the taxable gain and resets his reference level to the new cost basis even if the tax rate is zero. That is, he uses the gross cost view of setting his reference level as discussed in footnote 6.

investment depending solely on how it is traded. It might seem, therefore, that there is no benefit from diversification. However, this is not the case. For utility that is homogeneous of degree β as we have assumed, the value of investing wealth W in several assets in proportions w_i and subsequently reinvesting each portion without any rebalancing is $\sum_i (w_i W)^\beta v_i(1)$ where v_i is the separate valuation of asset i that depends on its mean and variance and the specific utility function. For $0 < \beta < 1$, utility is maximized with $w_i \propto [v_i(1)]^{1/(\beta-1)}$. Of course, this portfolio is generally not optimal absent the just-assumed no-rebalancing restriction, but it does demonstrate that investing in a single stock is a dominated policy so some diversification must be optimal. A more thorough analysis would show that when rebalancing is allowed, investors should deviate from a strictly constant Θ - θ policy whenever their allocation strays too far from the optimal.

However, before diversification and rebalancing can be studied completely, a more fundamental question must be answered: How are separate gains and losses aggregated under realization utility? Our model assumes complete narrow framing both across assets and over time. All utility bursts are evaluated separately and then their discounted values are summed over different assets and time periods. In a different model investors might aggregate contemporaneous gains and losses into a single utility burst and sum those over time. Both assumptions are psychologically plausible, but either creates problems in multi-period models, particularly continuous-time models.

If the utility function aggregates contemporaneous gains and losses, and the investor's utility is S-shaped, losses on different assets should be taken simultaneously if possible because marginal utility is decreasing in the magnitude of losses. This will lead to timing complications with investors postponing some loss taking while accelerating other losses to achieve synchronicity. On the other hand, gains on different assets should always be kept separate in time since their marginal utility is decreasing. This is trivial to achieve in continuous time with time-additive utility which so narrowly frames the time dimension that a separation of dt is sufficient for a completely independent evaluation. But is it reasonable to assume that two gains realized simultaneously have a different utility than the same two gains realized only instants apart? And if not, how long a time separation is required? Similar questions have arisen in the standard consumption-portfolio problem and various suggestions like recursive utility or Hindy, Huang, and Kreps (1992) intertemporal aggregation have been made. The question here is more difficult because consumption is naturally smoothed in intertemporal models, but here we have both smoothing of gains and lumping of losses to address.

7. Concluding Remarks

In this paper, we have built an intertemporal realization utility model to study investors' trading behavior. Highlighting the role reinvestment plays in a dynamic context, we have shown that investors may voluntarily sell at losses even though that has an immediate negative utility impact. The necessary condition for voluntary losses is either risk seeking behavior over at least some losses or decreasing scaling under which the magnitude of the utility of gains and losses realized with smaller reference levels is larger.

Under our model a two-point sales strategy is optimal; an investor sells for a gain when

the asset value rises to a fixed multiple of the reference level or sells at a loss at a fixed fraction of the reference level. We have provided a detailed calibration of the model showing that the trading data of Odean and others is in close agreement with such two-point strategies but is inconsistent with trading that is independent of the potential gain. We also introduced a modified form of Tversky-Kahneman utility which predicts, either alone or in a model of heterogeneous investors, the two trading points observed in the data.

We have discussed some properties predicting trading patterns and price effects. In particular, the model suggests a flattening of the capital market line and that idiosyncratic risk could have a negative risk premium. Both of these features are seen in the data. But other effects still need to be investigated. For example, how do trading patterns and volume evolve over time in different markets. What is the relation between realization utility and momentum?

There are several directions for future theoretical research. One important step is to study the diversification motive of realization-utility investors and solve a full portfolio problem. We do know that some diversification and rebalancing is optimal, but this means that the optimal sales strategy for a stock depends not only on its reference level but also on the prices and reference levels of the other assets held in the portfolio. In addition, assumptions must then be made about the proper reference level for a position that consists of shares purchased for different prices at different times.

It is also important to understand how realization utility interacts with other types of utility. Do investors also receive utility just from holding assets with paper gains even if they do not sell? Clearly investors also value consumption; are the motives to smooth consumption and to realize gains evaluated separately or combined somehow?

Finally if an S-shaped utility function is important, does probability weighting also have an effect on realization utility? It is a non-trivial task to incorporate probability weighting into an intertemporal setting because the law of iterated expectation does not hold if probabilities are replaced by decision weights. As shown in Barberis (2012) and Ingersoll (2012), cumulative probability weighting typically induces time inconsistency, and certain rules that define economic actions need to be imposed in order to further model this type of behavior.

Appendices

A: Notation

TABLE OF IMPORTANT NOTATION

\mathbb{E}	Expectations operator
G	Dollar size of a gain or a loss
k_s	Proportional transaction cost for selling stock
k_p	Proportional transaction cost for purchasing stock
K	Round-trip proportional transaction cost for selling the current asset being held and reinvesting in another asset, $K \equiv (1 - k_s)/(1 + k_p)$
\bar{n}	Average number of stocks held in a single investment account
\mathcal{O}	The Odean measure of the disposition effect, $\mathcal{O} \equiv PGR/PLR$
PGR	Proportion of gains realized
PLR	Proportion of losses realized
R	The reference level
Q_G	The probability that a given investment episode eventually ends with a gain
Q_L	The probability that a given investment episode eventually ends with a loss
S	The stock price
U	The utility burst function $U(G, R)$
u	The reduced form utility burst function $u(G/R) \equiv R^{-\beta}U(G, R)$
V	The undiscounted value function $V(X, R)$
v	The reduced form value function $v(x) \equiv R^{-\beta}V(X, R)$
X	The level of investment
x	Gross return relative to the reference level, $x \equiv X/R$
α_G	Parameter of risk aversion for gains in evaluating utility bursts
α_L	Parameter of risk seeking for losses in evaluating utility bursts
β	The scaling parameter gauging the impact of the reference level, R , on utility bursts
γ_1, γ_2	Characteristic roots of the partial differential equation, defined in equation (8)
δ	The subjective discount rate
η	Defined parameter $\eta \equiv 1 - 2\mu/\sigma^2$
Θ	Optimal sales multiple for a gain; sale occurs at $X = \Theta R$
θ	Optimal sales fraction for a loss; sale occurs at $X = \theta R$
κ	Parameter measuring the subjective effect of transaction costs in evaluating utility bursts
λ	Loss aversion parameter
μ	Growth rate of the stock price
ρ	Poisson intensity for random trading
σ	Logarithmic standard deviation of the stock price
σ_n	Standard deviation of number of stocks held in investment accounts
τ	Duration of an investment episode from purchase to sale
φ_G	The fraction of time an asset has an unrealized gain
φ_L	The fraction of time an asset has an unrealized loss

B: Verification of the Optimality of Constant Proportional Sales Policies

Consider the general sale and reinvestment problem given in the text. For any given realized stochastic price path and sales policy denote the original stock price and the stock prices at the points of sale as $S_0, S_1, \dots, S_n, \dots$. The dollar amount under investment is denoted as X_n^\pm where the minus and positive superscripts denote the value before and after paying the transaction costs at the n^{th} sale and repurchase. N_n denotes the number of shares purchased in the $(n-1)^{\text{st}}$ purchase and sold in the n^{th} sale. The relations among the X 's are given by the recurrences

$$X_n^+ = (1 - k_s)X_n^- / (1 + k_p) \equiv KX_n^- \quad X_{n+1}^- = X_n^+ S_{n+1} / S_n. \quad (\text{A1})$$

It is readily verified that

$$\begin{aligned} X_{n+j}^- &= N_n S_n (S_{n+1} / S_n) \cdots (S_{n+j} / S_{n+j-1}) K^{j-1} / (1 + k_p) = N_n S_{n+j} K^{j-1} / (1 + k_p) \\ X_{n+j}^+ &= KX_{n+j}^- = N_n S_{n+j} K^j / (1 + k_p). \end{aligned} \quad (\text{A2})$$

The reference level, R_n , established after the n^{th} sale and repurchase is proportional to both X_n^\pm . We have assumed that the investor considers both sets of transaction costs in setting his reference level with $R_n = X_n^+ = KX_n^-$. Alternatively he might ignore the repurchase costs setting $R_n = X_n^-(1 - k_s) = X_n^+(1 + k_p)$. To cover both of these and many other cases we define $R_n = K'X_n^+$. The subjective realized rate of return gain at the n^{th} sale is

$$\kappa X_n^- / R_{n-1} - 1 = (\kappa / K') X_n^- / X_{n-1}^+ - 1 = (\kappa / K') S_n / S_{n-1} - 1. \quad (\text{A3})$$

The utility burst at time t_n is

$$e^{-\delta t_n} R_{n-1}^\beta u(\kappa X_n^- / R_{n-1} - 1) \quad (\text{A4})$$

The argument of the utility function and the reference level are

$$\begin{aligned} \kappa X_n^- / R_{n-1} - 1 &= (\kappa / K') S_n / S_{n-1} - 1 \\ R_{n-1} &= K' X_{n-1}^+ = \frac{K' N_0 S_{n-1} K^{n-1}}{1 + k_p} = K' X_0^+ K^{n-1} \frac{S_{n-1}}{S_0} = K^{n-1} R_0 \frac{S_{n-1}}{S_0}. \end{aligned} \quad (\text{A5})$$

So for this particular realized outcome and strategy, lifetime future utility measured just after the n^{th} sale and repurchase is

$$\Upsilon_{t_n} = e^{-\delta t_n} R_n^\beta \sum_{j=1} e^{-\delta(t_{n+j} - t_n)} [K^{j-1} (S_{n+j-1} / S_n)]^\beta u((\kappa / K') S_{n+j} / S_{n+j-1} - 1). \quad (\text{A6})$$

Now consider a rule that generates the optimal sales policy. It sets gain and loss sales points for the first sale of $\bar{S}_1(S_0)$ and $\underline{S}_1(S_0)$ and contingent rules for the second sales points, $\bar{S}_2(S_1)$ and $\underline{S}_2(S_1)$, etc. These optimal policies cannot depend on time since the stochastic process

is time homogeneous and the investor has a constant subjective discount rate. Nor can the optimal policies depend on the reference level, R_n , (separately from the prevailing stock price, S_n) since looking ahead from time t_n , the reference level affects the realized utility only through the proportional scaling factor R_n^β . Finally, the optimal policies for the second sale can depend on S_0 only through S_1 since the former price doesn't affect the utility looking forward from time t_1 and the stock returns are independent and identically distributed. Therefore we must have $\bar{S}_2(S_1)/S_1 = \bar{S}_1(S_0)/S_0$, etc.

It is clear from this analysis that constant proportional transaction costs, a constant and proportional subjective interpretation of the realized return and reference level (κ and K'), a stock price process with independent and identically distributed returns, a constant rate of time preference, an infinite horizon, and a utility based on rates of return with power scaling factor are all necessary for a constant optimal policy.

C: Transversality Conditions

There are several different conditions required in our analysis to keep the value function finite or produce a well-defined optimal trading strategy. First the discount rate must be large enough or utility bursts far in the future dominate the value function and make its unbounded. This condition mimics the transversality conditions in the standard portfolio problem. Second the scaling parameter β must be not too large otherwise repeated selling lets total utility accumulate too quickly by increasing the reference level and thereby the utility bursts of future sales. Finally, transaction costs cannot be zero or the investor can repeatedly realize small gains with their very high marginal utility.

Suppose an investor adopts a constant sales policy, Θ , and never sells at a loss. With each sale the reference level increases by the factor $K\Theta$ so at the time of the n^{th} sale it will be $R_n = R_0(K\Theta)^{n-1}$. The n^{th} sale has a subjective gain of $\kappa\Theta - 1$ per dollar of the reference level or $(\kappa\Theta - 1)R_{n-1}$. The expected lifetime utility from a series of sales at gains and no sales at losses is

$$\mathbb{E} \sum_{n=1}^{\infty} e^{-\delta \tilde{t}_n} R_{n-1}^\beta u(\kappa\Theta - 1) = u(\kappa\Theta - 1) \frac{R_0^\beta}{(K\Theta)^\beta} \sum_{n=1}^{\infty} (K\Theta)^{n\beta} \mathbb{E}[e^{-\delta \tilde{t}_n}]. \quad (\text{A7})$$

Here \tilde{t}_n is the random time of the n^{th} sale. Note that the utility bursts realized are known; it is only the timing that is random. Since the times between successive sales are independent and identically distributed

$$\mathbb{E}[e^{-\delta \tilde{t}_n}] = \mathbb{E}[e^{-\delta \tilde{t}_1}] \mathbb{E}[e^{-\delta(\tilde{t}_2 - \tilde{t}_1)}] \dots \mathbb{E}[e^{-\delta(\tilde{t}_n - \tilde{t}_{n-1})}] = \left(\mathbb{E}[e^{-\delta \tilde{t}_1}] \right)^n, \quad (\text{A8})$$

and the final sum in (A7) is

$$\sum_{n=1}^{\infty} \left((K\Theta)^\beta \mathbb{E}[e^{-\delta \tilde{t}_1}] \right)^n \quad (\text{A9})$$

which converges if and only if $(K\Theta)^\beta \mathbb{E}[e^{-\delta \tilde{t}_1}] < 1$. The expected value in (A9) depends on Θ through the stopping time \tilde{t}_1 , but unless $\delta > 0$, it is at least one for any Θ so the sum is unbounded for any $\Theta > 1/K$ and infinite utility is possible. Therefore δ must be positive just as in the standard infinite-horizon investment problem.

The expectation in (A9) is a particular value of the Laplace transform³⁶ of the first-passage time density of the random variable, X_t , to ΘR ; that is,

$$\begin{aligned} \mathbb{E}[e^{-s\tilde{t}_1}] &= \exp\left(\frac{\ell n \Theta}{\sigma^2} \left[\mu - \frac{1}{2} \sigma^2 - \sqrt{(\mu - \frac{1}{2} \sigma^2)^2 + 2s\sigma^2} \right]\right) \\ \Rightarrow \mathbb{E}[e^{-\delta \tilde{t}_1}] &= \exp(-\gamma_1 \ell n \Theta) \end{aligned} \quad (\text{A10})$$

where γ_1 is the positive exponent in the solution of the differential equation for valuation as given in (8). Therefore, for the sum in (A9) to be bounded, we also require

$$1 > (K\Theta)^\beta \mathbb{E}[e^{-\delta \tilde{t}_1}] = K^\beta \Theta^{\beta - \gamma_1}. \quad (\text{A11})$$

The feasible policy choices are $\Theta > 1/\kappa \geq 1$, so infinite utility can be achieved if $\beta > \gamma_1$ with any feasible $\Theta > K^{-\beta/(\beta - \gamma_1)}$. If $0 \leq \beta \leq \gamma_1$, then the sum converges for all feasible policies.³⁷ Combining these results, necessary conditions for there to be no investment plans which lead to infinite expected utility are

$$\beta < \gamma_1 \quad \text{and} \quad \delta > 0. \quad (\text{A12})$$

These conditions depend only on the scaling of the burst utility function and are required whatever the functional form of the reduced utility $u(\cdot)$.

However, these conditions are not sufficient as unbounded utility might also be achieved, if the growth rate of the asset is too large. Suppose an investor adopts a gains-only policy ($\theta = 0$), then from (10) in the text, $C_2 = 0$. And as $\Theta \rightarrow \infty$, the initial value function is

$$v(1) = C_1 = \frac{u(\kappa\Theta - 1)}{\Theta^{\gamma_1} - (K\Theta)^\beta} \sim \begin{cases} \kappa^{\alpha_G} \Theta^{\alpha_G - \gamma_1} & \text{for scaled-TK} \\ \kappa^{\alpha_G} \Theta^{\alpha_G - \gamma_1} / \alpha_G & \text{for modified-TK} \end{cases} \quad (\text{A13})$$

since $\beta < \gamma_1$. This is clearly unbounded if $\alpha_G > \gamma_1$; therefore, a necessary and sufficient condition for no transversality violation is $\alpha_G \leq \gamma_1$ or in terms of the exogenous parameters

³⁶ The Laplace transform of a density function, $\mathcal{L}(f(t)) \equiv \mathbb{E}[e^{-st}]$, can be easily determined from the moment generating function, $\mathcal{M}(f(t)) \equiv \mathbb{E}[e^{st}]$, for a negative argument. The Laplace transform is defined for all values of μ (unlike the moment generating function) since the first passage time is a positive random variable.

³⁷ We have assumed that $\beta \geq 0$ to ensure participation; however, if $\beta < 0$, then the sum might diverge for $\Theta \approx 1$. The minimum feasible value for Θ is $1/\kappa$, so the convergence condition is $1 > K^\beta \kappa^{\gamma_1 - \beta}$. If the investor is subjectively fully cognizant of all transaction costs then $\kappa = K$, and this condition is met; otherwise for $\kappa < K$, the lower bound on β is $\gamma_1 \ell n(\kappa) / \ell n(\kappa/K)$.

$$\alpha_G[\mu + \frac{1}{2}(\alpha_G - 1)\sigma^2] \leq \delta. \quad (\text{A14})$$

This is similar to the transversality condition found in the standard infinite horizon portfolio problem. For linear over gains, μ cannot exceed the discount rate, but risk aversion expands the set of admissible values for μ .

Finally unbounded utility can be achieved if there are no transaction costs so there is no well-defined optimal strategy in the absence of costs. To demonstrate this, we construct a sequence of sales strategies and show that as the transaction costs go to zero, the constructed strategies lead to unbounded utility. Therefore, the maximizing strategy must also yield infinite utility in the limit of zero costs.

Consider a sequence of economies indexed by k . We assume that the sales and purchase costs are in the same proportion $c \geq 0$ ³⁸ for each step of the sequence, $k_s = k$ and $k_p = ck$. At each step the sales policies considered are

$$\Theta = 1 / (1 - k) + k^{\omega_0}, \quad \theta = 1 - k^{\omega_1}, \quad \text{where } 0 < \omega_0, \omega_1 < \frac{1}{2}. \quad (\text{A15})$$

These are not assumed to be the optimal strategies only feasible ones.

Initial utility is $v(1) = C_1 + C_2$ where these constants are defined in (10). Using a Taylor expansion for small k , the functions determining C_1 and C_2 are

$$\begin{aligned} c_1(\Theta) &= (\gamma_1 - \beta)k^{\omega_0} + \frac{1}{2}[\gamma_1(\gamma_1 - 1) - \beta(\beta - 1)]k^{2\omega_0} + o(k^{2\omega_0}) \\ c_1(\theta) &= (\beta - \gamma_1)k^{\omega_1} + \frac{1}{2}[\gamma_1(\gamma_1 - 1) - \beta(\beta - 1)]k^{2\omega_1} + o(k^{2\omega_1}) \\ c_2(\Theta) &= (\gamma_2 - \beta)k^{\omega_0} + \frac{1}{2}[\gamma_2(\gamma_2 - 1) - \beta(\beta - 1)]k^{2\omega_0} + o(k^{2\omega_0}) \\ c_2(\theta) &= (\beta - \gamma_2)k^{\omega_1} + \frac{1}{2}[\gamma_2(\gamma_2 - 1) - \beta(\beta - 1)]k^{2\omega_1} + o(k^{2\omega_1}). \end{aligned} \quad (\text{A16})$$

The constants are

$$\begin{aligned} C_1 &= \frac{(\beta - \gamma_2)k^{\omega_1 + \alpha_G \omega_0} - \lambda(\beta - \gamma_2)k^{\omega_0 + \alpha_L \omega_1} + o(k^{\min(\omega_0 + \alpha_L \omega_1, \omega_1 + \alpha_G \omega_0)})}{C(k^{2\omega_0 + \omega_1} + k^{2\omega_1 + \omega_0}) + o(k^{\min(2\omega_0 + \omega_1, 2\omega_1 + \omega_0)})} \\ C_2 &= \frac{(\gamma_1 - \beta)k^{\omega_1 + \alpha_G \omega_0} - \lambda(\gamma_1 - \beta)k^{\omega_0 + \alpha_L \omega_1} + o(k^{\min(\omega_0 + \alpha_L \omega_1, \omega_1 + \alpha_G \omega_0)})}{C(k^{2\omega_0 + \omega_1} + k^{2\omega_1 + \omega_0}) + o(k^{\min(2\omega_0 + \omega_1, 2\omega_1 + \omega_0)})} \end{aligned} \quad (\text{A17})$$

where

$$\begin{aligned} C &= \frac{1}{2}(\gamma_2 - \gamma_1)[\gamma_1\gamma_2 - \frac{1}{2}\beta(\gamma_1 + \gamma_2) + \frac{1}{2}\beta^2] \\ &\geq \frac{1}{2}(\gamma_2 - \gamma_1)\left(\gamma_1\gamma_2 - \frac{1}{2}\beta\gamma_1 - \beta\gamma_2 + \frac{1}{2}\beta^2\right) = \frac{1}{2}(\gamma_1 - \gamma_2)(\gamma_1 - \beta)\left(\frac{1}{2}\beta - \gamma_2\right) > 0. \end{aligned}$$

³⁸ Given that ω_0 and ω_1 are both less than $\frac{1}{2}$, all the terms associated with c are included in the higher-order terms $o(k^{2\omega_0})$ and $o(k^{2\omega_1})$.

If we choose the converging rates, ω_0 and ω_1 , such that $\omega_0/\omega_1 > (1-\alpha_L)/(1-\alpha_G)$, then from (A17) it is easy to verify that $C_1, C_2 \rightarrow \infty$, as $k \rightarrow 0$. So this strategy gives infinite utility in the limit and the truly optimal must as well.

D: Proof of Proposition 1

As illustrated in Figure 2 there can be two local maxima to our optimization problem. The one-point maximum is a corner solution with $\theta = 0$; the two-point maximum is an interior maximum with $\theta > 0$. Which local maxima is the global maximum depends on the specific economic and utility parameters, but it can be most easily characterized by the loss aversion parameter, λ . The value function for the one-point maximum does not depend on λ since no losses are ever realized. The value function for the two-point maximum is obviously decreasing in λ . Therefore, there is a single critical value of λ at which the two value functions are equal that marks the change in regime.

Denote the two-point and one-point value functions as $v^{(2)}(x) = C_1^{(2)}x^{\gamma_1} + C_2^{(2)}x^{\gamma_2}$ and $v^{(1)}(x) = C_1^{(1)}x^{\gamma_1}$. Since, for the critical value of λ_* , the value functions are equal everywhere that they are defined, we must have $C_1^{(2)} = C_1^{(1)} = C_1$ and $C_2^{(2)} = 0$. From (9), the boundary conditions are

$$C_1\Theta^{\gamma_1} = u(\kappa\Theta - 1) + (K\Theta)^\beta C_1 \quad C_1\theta^{\gamma_1} = u(\kappa\theta - 1) + (K\theta)^\beta C_1 \quad (\text{A18})$$

where $u(g) \equiv U(g, 1)$ is the reduced-form version of the utility burst for the proportional gain or loss $g \equiv G/L$. The smooth-pasting conditions for a maximum are

$$\gamma_1 C_1 \Theta^{\gamma_1-1} = \kappa u'(\kappa\Theta - 1) + \beta K^\beta \Theta^{\beta-1} C_1 \quad \gamma_1 C_1 \theta^{\gamma_1-1} = \kappa u'(\kappa\theta - 1) + \beta K^\beta \theta^{\beta-1} C_1. \quad (\text{A19})$$

Each boundary condition in (A18) yields a formula for C_1 which must give the same value

$$C_1 = \frac{u(\kappa\Theta - 1)}{\Theta^{\gamma_1} - (K\Theta)^\beta} = \frac{u(\kappa\theta - 1)}{\theta^{\gamma_1} - (K\theta)^\beta}. \quad (\text{A20})$$

Similarly, from (A19)

$$C_1 = \frac{\kappa u'(\kappa\Theta - 1)}{\gamma_1 \Theta^{\gamma_1-1} - \beta K^\beta \Theta^{\beta-1}} \quad \text{and} \quad C_1 = \frac{\kappa u'(\kappa\theta - 1)}{\gamma_1 \theta^{\gamma_1-1} - \beta K^\beta \theta^{\beta-1}} \quad (\text{A21})$$

Substituting the explicit utility function into (A20) and solving for λ_* gives

$$\lambda_* = - \frac{(\kappa\Theta - 1)^{\alpha_G} \theta^{\gamma_1} - K^\beta \theta^\beta}{(1 - \kappa\theta)^{\alpha_L} \Theta^{\gamma_1} - K^\beta \Theta^\beta}. \quad (\text{A22})$$

Similarly the two equations in (A21) give

$$\begin{aligned}
0 &= (\alpha_G - \gamma_1)\kappa\Theta^{\gamma_1+1-\beta} + \gamma_1\Theta^{\gamma_1-\beta} - (\alpha_G - \beta)K^\beta\kappa\Theta - \beta K^\beta \\
0 &= (\alpha_L - \gamma_1)\kappa\theta^{\gamma_1+1-\beta} + \gamma_1\theta^{\gamma_1-\beta} - (\alpha_L - \beta)K^\beta\kappa\theta - \beta K^\beta.
\end{aligned} \tag{A23}$$

These equations can be re-expressed as

$$\Theta^{\gamma_1} - (K\Theta)^\beta = \frac{(K\Theta)^\beta(\gamma_1 - \beta)(\kappa\Theta - 1)}{(\alpha_G - \gamma_1)\kappa\Theta + \gamma_1} \quad \theta^{\gamma_1} - (K\theta)^\beta = \frac{(K\theta)^\beta(\gamma_1 - \beta)(\kappa\theta - 1)}{(\alpha_L - \gamma_1)\kappa\theta + \gamma_1}. \tag{A24}$$

Substituting back into (A22) gives

$$\lambda_* = \frac{(\kappa\Theta - 1)^{\alpha_G - 1}\theta^\beta}{(1 - \kappa\theta)^{\alpha_L - 1}\Theta^\beta} \times \frac{(\alpha_G - \gamma_1)\kappa\Theta + \gamma_1}{(\alpha_L - \gamma_1)\kappa\theta + \gamma_1} \tag{A25}$$

which is the desired expression in (12).

Correspondingly for the modified-TK utility introduced in the calibration session, (A20) now gives

$$\lambda_* = \frac{\alpha_L}{\alpha_G} \frac{(\kappa\Theta)^{\alpha_G} - 1}{(\kappa\theta)^{\alpha_L} - 1} \frac{\theta^{\gamma_1} - K^\beta\theta^\beta}{\Theta^{\gamma_1} - K^\beta\Theta^\beta}, \tag{A26}$$

instead of (A22), and (A21) now gives

$$\begin{aligned}
0 &= (\alpha - \gamma_1)\kappa^{\alpha_G}\Theta^{\gamma_1+\alpha_G-\beta} + \gamma_1\Theta^{\gamma_1-\beta} - (\alpha_G - \beta)K^\beta(\kappa\Theta)^{\alpha_G} - \beta K^\beta \\
0 &= (\alpha_L - \gamma_1)\kappa^{\alpha_L}\theta^{\gamma_1+\alpha_L-\beta} + \gamma_1\theta^{\gamma_1-\beta} - (\alpha_L - \beta)K^\beta(\kappa\theta)^{\alpha_L} - \beta K^\beta.
\end{aligned} \tag{A27}$$

Together (A26) and (A27) give

$$\lambda_* = \frac{\alpha_L}{\alpha_G} \left(\frac{\theta}{\Theta} \right)^\beta \frac{(\alpha_G - \gamma_1)\kappa^{\alpha_G}\Theta^{\alpha_G} + \gamma_1}{(\alpha_L - \gamma_1)\kappa^{\alpha_L}\theta^{\alpha_L} + \gamma_1} \tag{A28}$$

in place of (A25).

As $\beta \rightarrow \gamma_1$ from below, both (A23) and (A27) yield $\theta \approx K^{\beta/(\gamma_1-\beta)} \rightarrow 0$. Substituting this into either (A25) or (A28) gives $\lambda_* \rightarrow 0$, which implies that in this limiting case, voluntary loss realization will never take place. ■

E: Proofs of Proposition 2 to 6

Proof of Proposition 2: The sales times are first-passage times for the lognormal diffusion which are known to have the densities

$$\begin{aligned}
q_G(\tau) &= \sum_{n=-\infty}^{\infty} \frac{\ell n \Theta + z'_n}{\sigma \tau^{3/2}} \exp(\frac{1}{2} \eta z'_n) \phi\left(\frac{\ell n \Theta + z'_n - (\mu - \frac{1}{2} \sigma^2) \tau}{\sigma \sqrt{\tau}}\right) \\
q_L(\tau) &= \sum_{n=-\infty}^{\infty} \frac{-\ell n \theta - z'_n}{\sigma \tau^{3/2}} \exp(\frac{1}{2} \eta z'_n) \phi\left(\frac{\ell n \theta + z'_n - (\mu - \frac{1}{2} \sigma^2) \tau}{\sigma \sqrt{\tau}}\right)
\end{aligned} \tag{A29}$$

$$\text{where } \eta \equiv 1 - 2\mu/\sigma^2 \quad z'_n \equiv 2n\ell n(\Theta/\theta)$$

and $\phi(\cdot)$ is the standard normal density. The ultimate resolution probabilities, Q_G and Q_L , and the expected duration, $\mathbb{E}[\tau]$, can be determined by integrating q_G and q_L .³⁹ Since the asset returns are independent and identically distributed, a sequence of investments forms a renewal process, and the inverse relation between the expected duration and the expected number of episodes is given by the Elementary Renewal Theorem of Statistics.

The stochastic process for $x \equiv X/R$ is a Markov process with bounded support. It is a diffusion everywhere except at $x=1$ and the two boundaries, θ and Θ . It is not a diffusion at those points because whenever x reaches either boundary, it returns immediately to $x=1$ due to the sale and reinvestment. The steady-state distribution of x is the solution to the Kolmogorov forward equation

$$\begin{aligned}
0 &= \frac{1}{2} \frac{d^2}{dx^2} [\sigma^2 x^2 f] - \frac{d}{dx} [\mu x f] = \frac{1}{2} \sigma^2 x^2 f_{xx} + (2\sigma^2 - \mu) x f_x + (\sigma^2 - \mu) f \\
&\text{for } x \in (\theta, 1) \cup (1, \Theta) \quad \text{and subject to } f(\theta) = f(\Theta) = 0.
\end{aligned} \tag{A30}$$

The density must vanish at the boundaries just as it does with an absorbing barrier because the infinite variation in the diffusion process assures that the barrier will be reached with probability approaching unity whenever the diffusion enters a small neighborhood of the barrier.

Since x is not a diffusion at 1, the differential equation does not hold at that point, and the equation must be solved separately in the two regions then pieced together. The general solution to the differential equation is $C_1 x^{-1} + C_2 x^{-\eta}$ where $\eta \equiv 1 - 2\mu/\sigma^2$, and the two constants differ in the two regions. The two boundary conditions, the continuity of the density at $x=1$, and the unit mass of f over the entire region supply the four equations needed to determine the four constants. The density and cumulative distribution function are

³⁹ See for example, Cox and Miller (1965) for the densities in (A29). The ultimate resolution probabilities in (13) can also be determined directly as the solutions to the time-independent backward equation $0 = \frac{1}{2} \sigma^2 x^2 q'' + \mu x q'$. Solutions to this equation give the time-invariant probability of any event measurable in terms of the current variable x . Q_G is the value $q(1)$ from the solution subject to $q(\theta) = 0$ and $q(\Theta) = 1$ while Q_L is the value from the solution with the boundary conditions reversed. The expected remaining holding time until a sale, $T(x)$, is also a function of the current location. By the law of iterated expectations, it satisfies $T(x) = \mathbb{E}[T(x+dx)] + dt$. Using Itô's lemma, T is therefore the solution to $0 = \frac{1}{2} \sigma^2 x^2 T'' + \mu x T' + 1$ subject to $T(\theta) = T(\Theta) = 0$. The initial expected holding period is $\mathbb{E}[\tau] = T(1)$ which is given in (15).

$$f(x) = \begin{cases} \frac{(\Theta^n - 1)(\theta^n x^{-n} - 1)x^{-1}}{(1 - \theta^n)\ell n \Theta + (\Theta^n - 1)\ell n \theta} \\ \frac{(\theta^n - 1)(\Theta^n x^{-n} - 1)x^{-1}}{(1 - \theta^n)\ell n \Theta + (\Theta^n - 1)\ell n \theta} \end{cases} \quad \text{and} \quad F(x) = \begin{cases} \frac{(\Theta^n - 1)(\ell n(\theta/x) + \frac{1}{n}[1 - (\theta/x)^n])}{(1 - \theta^n)\ell n \Theta + (\Theta^n - 1)\ell n \theta} & x \leq 1 \\ 1 - \frac{(\theta^n - 1)(\ell n(x/\Theta) + \frac{1}{n}[(\Theta/x)^n - 1])}{(1 - \theta^n)\ell n \Theta + (\Theta^n - 1)\ell n \theta} & x \geq 1. \end{cases} \quad (\text{A31})$$

The probabilities in equation (14) are $\varphi_L = 1 - \varphi_G = F(1)$. ■

Proof of Proposition 3: It is well-known that the duration of each episode of a Poisson process has an exponential distribution. The cumulative probability that the sale occurs after a holding period of τ is $e^{-\rho\tau}$, and the average duration is $1/\rho$. At the end of an episode of lasting τ , $z \equiv \ell n x$ has a normal distribution with mean $(\mu - \sigma^2/2)\tau$ and variance $\sigma^2\tau$. Therefore, the expected price ratios conditional on a sale at a gain and a loss are

$$\begin{aligned} \bar{\Theta} &= (Q_G)^{-1} \int_0^\infty \rho e^{-\rho\tau} \int_0^\infty \frac{1}{\sigma\sqrt{\tau}} e^z \phi\left(\frac{z - (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right) dz d\tau = -\frac{\rho(1 - \Psi^-)}{(\rho - \mu)\Psi^-} \\ \bar{\theta} &= (Q_L)^{-1} \int_0^\infty \rho e^{-\rho\tau} \int_{-\infty}^0 \frac{1}{\sigma\sqrt{\tau}} e^z \phi\left(\frac{z - (\mu - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right) dz d\tau = \frac{\rho(\Psi^+ - 1)}{(\rho - \mu)\Psi^+} \end{aligned} \quad (\text{A32})$$

with $\rho > \mu$ required for $\bar{\Theta}$ to be finite.

The probability, $H(x)$, that a given episode eventually events with a loss conditional on the current asset value ratio, $x = X/R$, satisfies the modified backward equation

$$0 = \frac{1}{2}\sigma^2 x^2 H'' + \mu x H' + \rho(\mathbf{1}_{x < 1} - H) \quad \text{subject to } H(0) = 1, H(\infty) = 0 \quad (\text{A33})$$

where $\mathbf{1}_{x < 1}$ is the indicator function that the asset currently has a paper loss. This final term is the probability per unit time that H changes discontinuously to one (zero) if a sales event occurs when x is less (greater) than 1 resulting in a realized loss (gain). Solving this equation in the two regions and matching it and its derivative at the boundary $x = 1$ gives

$$H(x) = \begin{cases} \frac{\Psi^+}{\Psi^+ - \Psi^-} x^{\Psi^-} & x > 1 \\ 1 + \frac{\Psi^-}{\Psi^+ - \Psi^-} x^{\Psi^+} & x < 1. \end{cases} \quad (\text{A34})$$

The probabilities that a new investment episode ends in a sale at a loss and gain are $Q_L = H(1)$ and $Q_G = 1 - H(1)$. Since the Poisson events determining sales are independent of the stochastic process of the stock price, the probability of realizing a loss for any episode must equal the steady-state probability of holding an unrealized paper loss; that is, $\varphi_L = Q_L$. Similarly for gains, $\varphi_G = Q_G = 1 - H(1)$. ■

Proof of Proposition 4: In a given sample of T observed sales, the observed proportion of gains realized is⁴⁰

$$PGR = \frac{\sum_{t=1}^T \tilde{s}_t^G}{\sum_{t=1}^T (\tilde{s}_t^G + \sum_{i=1, i \neq i_t}^{\tilde{n}_t} \tilde{u}_{ii}^G)}$$

$$\tilde{s}_t^G = \begin{cases} 1 & \text{if the } t^{\text{th}} \text{ trade is a realized gain} \\ 0 & \text{if the } t^{\text{th}} \text{ trade is a realized loss} \end{cases}$$

$$\tilde{u}_{ii}^G = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ stock held in the account executing the } t^{\text{th}} \text{ trade has a paper gain} \\ 0 & \text{if the } i^{\text{th}} \text{ stock held in the account executing the } t^{\text{th}} \text{ trade has a paper loss} \end{cases} \quad (\text{A35})$$

\tilde{n}_t = the number of stocks held in the account making the t^{th} trade
and i_t is the index of the stock sold in the t^{th} trade.

Since each stock has an equal probability of being sold in the steady state, by the Weak Law of Large Numbers, the fraction of trades made by accounts holding n shares converges to $n\pi_n/\bar{n}$ where π_n is the fraction of investors who have n stocks in their accounts, and

$$\text{plim } T^{-1} \sum_{t=1}^T \tilde{s}_t^G = Q_G$$

$$\text{plim } T^{-1} \sum_{t=1}^T \sum_{i=1, i \neq i_t}^{\tilde{n}_t} \tilde{u}_{ii}^G = \sum_n (\pi_n n / \bar{n}) (n-1) \varphi_G = (\bar{n} + \sigma_n^2 / \bar{n} - 1) \varphi_G . \quad (\text{A36})$$

From Slutsky's Theorem, that the probability limit of a function of random variables is equal to the function of the separate probability limits, the plim of PGR is the ratio of these two quantities. Similar reasoning derives the plim of PLR . Equation (18) follows from an additional application of Slutsky's Theorem to the ratio, PGR/PLR . ■

Proof of Proposition 5: Because asset returns are independent and identically distributed, a sequence of investment episodes for any investor is a renewal process. Therefore, by the Elementary Renewal Theorem the average number of investment episodes per unit time in a single sequence of trades for a type i investor is $1/\mathbb{E}[\tau_i]$. Since type i investors hold the fraction $\pi_i n_i$ of the stocks, their trades form the fraction $\pi_i n_i / \mathbb{E}[\tau_i]$ of all trades on average, and this weight is used to compute the average statistics. In determining $\bar{\varphi}_G$, this weight is applied to the sum of the probabilities that each of the $n_i - 1$ other stocks not traded by the given type i investor has a paper gain. The probability limit then follows from Slutsky's Theorem. The same reasoning applies to PGR and PLR . ■

Proof of Proposition 6: This proof is similar to that for Proposition 5. Here the stocks in group i form the fraction $n_i / \mathbb{E}[\tau_i]$ of all trades on average, and this weight is used to compute the average statistics. In determining $\bar{\varphi}_G$, this weight is applied to the sum of probabilities that each of the $N - 1$ stocks not traded has a paper gain when one of the n_i stocks is traded. The probability limit

⁴⁰ Consistent with the empirical implementation of Odean (1998), the calculation in (A35) counts paper gains and losses only when at least one stock is sold in an account. Also (A35) assumes that each trade involves the sale of a single asset in a given account. In a continuous-time model, two less than perfectly correlated assets will never (with probability one) reach their sales thresholds simultaneously.

then follows from Slutsky's Theorem. The same reasoning applies to *PGR* and *PLR*. ■

F: Proof of Proposition 7

To prove this proposition, we use the following lemma.

Lemma: The three conditions (ii), (iii), and (iv) in Proposition 7

$$(i) \delta > 0 \quad (ii) \frac{\partial U}{\partial G} > 0 \quad (iii) \frac{\partial^2 U}{\partial G^2} \leq 0 \quad (iv) \frac{\partial^2 U}{\partial G \partial R} \geq 0. \quad (A37)$$

lead to the two relations

$$U(X - R, R) \leq \int_R^X \frac{\partial U(0^-, r)}{\partial G} dr \quad \text{if } X < R, \quad U(X - R, R) \leq \int_R^X \frac{\partial U(0^+, r)}{\partial G} dr \quad \text{if } X > R, \quad (A38)$$

Proof: For $X > R$, we have

$$U(X - R, R) = \int_R^X \frac{\partial U(r - R, R)}{\partial G} dr \leq \int_R^X \frac{\partial U(0^+, R)}{\partial G} dr \leq \int_R^X \frac{\partial U(0^+, r)}{\partial G} dr. \quad (A39)$$

The first inequality holds because marginal utility is decreasing (assumption iii), and the second integrand is evaluated at 0^+ , its smallest argument in the range of the integral. The second inequality holds because marginal utility is increasing in R (assumption iv) and each of the reference levels at which the integrand is evaluated are above R . A similar proof holds when $X < R$. ■

Proof of Proposition 7: Since there are no transaction costs, the reference level is constant between sales, and the reference level is updated to the investment value after a sale, the reference level for the k^{th} sale is $R_k = X_{k-1}$. For any given realized stochastic path of X , assume there is a realized loss, and denote its reference level as $R_I = X_{I-1}$. By assumption, $X_I < X_{I-1}$. First, assume that the asset is eventually sold at a price above the reference level for this loss. Let J denote the first such sale; that is, for some J , $X_J > X_{I-1} > X_{J-1}$. The total utility realized from the sales I through J is

$$\begin{aligned} \sum_{k=I}^J e^{-\delta t_k} U(X_k - X_{k-1}, X_{k-1}) &= \sum_{k=I}^{J-1} e^{-\delta t_k} U(X_k - X_{k-1}, X_{k-1}) + e^{-\delta t_J} U(X_{I-1} - X_{J-1}, X_{J-1}) \\ &\quad + e^{-\delta t_J} U(X_J - X_{J-1}, X_{J-1}) - e^{-\delta t_J} U(X_{I-1} - X_{J-1}, X_{J-1}). \end{aligned} \quad (A40)$$

The equality is a tautology with the utility from a fictitious sale occurring at time t_J at value X_{I-1} both added and subtracted. This lets us split the sales into two parts. The first line includes sales for which the gains and losses (though not their utility) net exactly to zero. The utility of this part is

$$\begin{aligned}
& \sum_{k=I}^{J-1} e^{-\delta t_k} U(X_k - X_{k-1}, X_{k-1}) + e^{-\delta t_J} U(X_{I-1} - X_{J-1}, X_{J-1}) \\
& \leq \sum_{\substack{k=I \\ X_k < X_{k-1}}}^{J-1} e^{-\delta t_k} \int_{X_{k-1}}^{X_k} \frac{\partial U(0^-, x)}{\partial G} dx + \sum_{\substack{k=I \\ X_k > X_{k-1}}}^{J-1} e^{-\delta t_k} \int_{X_{k-1}}^{X_k} \frac{\partial U(0^+, x)}{\partial G} dx \\
& \quad + e^{-\delta t_J} \int_{X_{J-1}}^{X_{I-1}} \frac{\partial U(0^+, x)}{\partial G} dx < 0.
\end{aligned} \tag{A41}$$

This contribution to utility is negative since each element of the first integrand over losses has a matching element in the second or third integrand of gains. Note that the integrals in the first sum are negative since the lower limits are larger than the upper limits and the integrand is strictly positive by assumption (ii). Assumptions (i) through (iv) assure that the marginal utility of each loss element is greater than that of the matching gain element. In particular, marginal utility is decreasing in G (assumption iii) so evaluating the marginal utilities of gains at 0^+ makes them smaller than the marginal utilities of losses evaluated at 0^- . Marginal utility is also decreasing in time (assumption i and ii) and increasing in R (assumption iv), and each gain occurs later and has a smaller reference level than the corresponding loss.⁴¹

Since the terms in the first line of (A40) are negative, the total utility realized from the sales I through J is less than the sum of the two terms in the second line

$$\begin{aligned}
& \sum_{k=I}^J e^{-\delta t_k} U(X_k - X_{k-1}, X_{k-1}) < e^{-\delta t_J} U(X_J - X_{J-1}, X_{J-1}) - e^{-\delta t_J} U(X_{I-1} - X_{J-1}, X_{J-1}) \\
& = e^{-\delta t_J} \int_{X_{I-1}}^{X_J} \frac{\partial U(x - X_{J-1}, X_{J-1})}{\partial G} dx \leq e^{-\delta t_J} \int_{X_{I-1}}^{X_J} \frac{\partial U(x - X_{I-1}, X_{I-1})}{\partial G} dx \\
& = e^{-\delta t_J} U(X_J - X_{I-1}, X_{I-1}).
\end{aligned} \tag{A42}$$

The new inequality in the second line holds because of assumptions (iii) and (iv). The final comparison shows that holding the stock with reference level X_{I-1} and selling at X_J lead to strictly higher utility than that provided by the original plan, which, therefore, cannot be optimal.

Now suppose there is no later sale that occurs at a level above X_{I-1} . In this case, the accumulated utility up through every sale J after I is weakly negative. Precisely, using the same reasoning as before,

$$\sum_{k=I}^J e^{-\delta t_k} U(X_k - X_{k-1}, X_{k-1}) < \sum_{\substack{k=I \\ X_k < X_{k-1}}}^J e^{-\delta t_k} \int_{X_{k-1}}^{X_k} \frac{\partial U(0^-, x)}{\partial G} dx + \sum_{\substack{k=I \\ X_k > X_{k-1}}}^J e^{-\delta t_k} \int_{X_{k-1}}^{X_k} \frac{\partial U(0^+, x)}{\partial G} dx < 0. \tag{A43}$$

⁴¹ The final inequality is strict because condition (i) on δ is also strict. A strict inequality for (iii) or (iv) would also serve to make voluntary losses strictly sub-optimal.

This is true even for an infinite investment horizon with $J \rightarrow \infty$. Therefore, holding the stock at a reference level of X_{t-1} with no subsequent sales is strictly better than the original plan. ■

References

- Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance* 61, 259-299.
- Barber, Brad, and Terrance Odean. 2000. "Trading is Hazardous to Your Wealth: the Common Stock Performance of Individual Investors." *Journal of Finance* 55, 773-806.
- Barberis, Nicholas. 2012. "A Model of Casino Gambling." *Management Science* 58, 35-51.
- Barberis, Nicholas, and Wei Xiong. 2009. "What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation" *Journal of Finance* 64, 751-784.
- Barberis, Nicholas, and Wei Xiong. 2012. "Realization Utility." *Journal of Financial Economics* 104, 251-271.
- Ben-David, Itzhak, and David Hirshleifer. 2012. "Are Investors Really Reluctant to Realize their Losses? Trading Responses to Past Returns and the Disposition Effect" *Review of Financial Studies*, forthcoming.
- Cox, David R., and Hilton D. Miller. 1965. *The Theory of Stochastic Processes* (Wiley, New York)
- Frydman, Cary, Nicholas Barberis, Colin Camerer, Peter Bossaerts, and Antonio Rangel. 2012. "Testing Theories of Investor Behavior Using Neural Data." working paper, California Institute of Technology.
- Goetzmann, William, and Alok Kumar. 2008. "Equity Portfolio Diversification." *Review of Finance* 12, 433-463.
- Grinblatt, Mark, and Bing Han. 2005. "Prospect Theory, Mental Accounting, and Momentum." *Journal of Financial Economics* 78, 311-339.
- Hens, Thorsten, and Martin Vlcek. 2011. "Does Prospect Theory Explain the Disposition Effect?" *Journal of Behavioral Finance* 12, 141-157.
- Hindy, Ayman, Chi-Fu Huang, and David Kreps. 1992. "On Intertemporal Preferences in Continuous Time: The Case of Certainty." *Journal of Mathematical Economics* 21, 401-440.
- Ingersoll, Jonathan E. 2012. "Cumulative Prospect Theory Portfolios and the Representative Investor" working paper, Yale University.
- Ingersoll, Jonathan E., and Lawrence J. Jin. 2012. "Realization Utility, the Disposition Effect and Trading Patterns" working paper, Yale University.

- Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, 263-292.
- Kaustia, Markku. 2010. "Prospect Theory and the Disposition Effect" *Journal of Financial and Quantitative Analysis* 45, 791-812.
- Kyle, Albert, Hui Ou-Yang, and Wei Xiong. 2006. "Prospect Theory and Liquidation Decisions." *Journal of Economic Theory* 129, 273-288.
- Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53, 1775-1798.
- Odean, Terrance. 1999. "Do Investors Trade Too Much?" *American Economic Review* 89, 1279-1298.
- Prelec, Drazen. 1998. "The Probability Weighting Function" *Econometrica* 66, 497-527.
- Shefrin, Hersh, and Meir Statman. 1985. "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence." *Journal of Finance* 40, 777-790.
- Thaler, Richard. 1999. "Mental Accounting Matters." *Journal of Behavioral Decision-Making* 12, 183-206.
- Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5, 297-323.
- Weber, Martin, and Colin Camerer. 1998. "The Disposition Effect in Security Trading: An Experimental Analysis" *Journal of Economic Behavior & Organization* 33, 167-184.
- Wu, George, and Richard Gonzalez. 1996. "Curvature of the Probability Weighting Function." *Management Science* 42, 1676-1690.