Dynamic Economic Decision Problems under Behavioural Preferences and Market Imperfections

by

Alex Sing-lam Tse

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

December 2016

Contents

List of Tables

List of Figures

Acknowledgments

I must first express my utmost gratitude to my supervisors David Hobson and Vicky Henderson for their excellent guidance during my PhD study. Apart from providing me a lot of insightful advices which polish my academic work, their support and encouragement over the last couple of years have been the invaluable input into my professional development.

I also wish to acknowledge the financial support from a Chancellor's International Scholarship offered by Warwick which made my study here possible.

Care from friends and family has always been important to me. My special thanks go to Candy for her love, patience and flexibility during my course of study.

Last but not least, I would like to thank Mike K.P. So for igniting my passion in academic research in the first place. Without his mentorship back in year 2007, my path of life would have realised in a very different way.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. Work based on collaborative research is declared as follows:

Chapter 1 is a joint work with Vicky Henderson and David Hobson based on the paper "Can Probability Weighting Help Prospect Theory Explain the Disposition Effect?", available on SSRN 2823449.

Chapter 2 is a joint work with Vicky Henderson and David Hobson based on the paper "Randomized Strategies and Prospect Theory in a Dynamic Context", available on SSRN 2531457.

Chapter 3 is a joint work with David Hobson and Yeqi Zhu based on the paper "A Multi-asset Investment and Consumption Problem with Transaction Costs", which is an extensive generalisation of an older version of the work "Multi-asset Investment and Consumption Problem with Infinite Transaction Costs" (arXiv:1409.8037) authored by David Hobson and Yeqi Zhu.

Abstract

This thesis is a collection of three individual works on dynamic economic decision problems which go beyond expected utility maximisation in complete markets. The first chapter introduces an asset liquidation model under prospect theory preferences. We demonstrate that the probability weighting component of the model can predict liquidation strategies which better fit the empirical patterns of investors' stock trading behaviours, when compared to models which do not incorporate probability weighting. The second chapter explores the role of randomised strategies in an exit-timing problem faced by a prospect theory agent. Several new insights are offered: in a discrete model, access to randomisation can strictly improve the economic value to the agent; in a continuous time counterpart, allowing randomisation will significantly alter the prediction of an agent's behaviours and more realistic exit-strategies would be observed in contrast to the results from the existing literature. The final chapter studies an extension to the Merton's optimal investment and consumption problem under transaction costs, where the agent can also dynamically invest in a liquid hedging asset without a trading fee. We provide a complete solution. Important properties of the problem such as well-posedness conditions and comparative static results are derived.

Preface: an overview of the thesis

In a typical mathematical model of investment, it is commonly assumed that the underlying agent is an expected utility maximiser with trading access to a frictionless market. Whilst the assumptions of an expected utility criteria and a perfect market help simplify the analysis, these specifications are not always good descriptions of the real world. On the one hand, the paradigm of expected utility hypothesis is under constant challenges by both designed experiments in laboratories and empirical anomalies in financial markets. On the other hand, trading can be costly in view of transaction fees. It may thus be inefficient to implement the optimal strategies advocated by standard economic models which often involve continuous portfolio rebalancing.

This thesis attempts to expand the classical theory by studying three different dynamic economic decision problems featuring behavioural preferences and market imperfections. The unifying goal is to investigate whether the extra features introduced can better reconcile the model predictions and real world phenomena. Ultimately, the results documented in this thesis will be useful for advancing our understanding to individuals' decisions in different economic contexts such as stock trading, exit-timing of casino gambling and portfolio choice. The first two chapters share a common theme involving an optimal stopping problem under prospect theory (PT), arguably the most popular behavioural model of decision making under risk proposed by Tversky and Kahneman (1992). The final chapter considers a separate topic on market friction where trading in a certain asset class incurs a proportional transaction cost. While each chapter is structured in a rather self-contained manner and thus can be read in any order in principle, readers might find Chapter 1 a useful prerequisite to Chapter 2.

In Chapter 1, we investigate the implications of PT preferences in an asset liquidation model. In the literature of empirical finance, the price disposition effect is a well-documented anomaly which refers to the tendency that investors hold losing stocks for too long but sell winning stocks too early. However, standard models of asset sale often fail to calibrate the strength of the price disposition effect satisfactorily. Very often they either do not predict voluntary sale at losses, or the price disposition effect implied is too extreme in comparison to the empirical data. We focus on the probability weighting component of PT which is typically omitted in the existing literature. On the theoretical side, we apply and extend the recent mathematical results regarding optimal stopping with probability weighting. A few extra but mild sufficient conditions on the agent's preference functions are provided which can lead to an optimal trading strategy with simple structure. We then explain with theoretical justifications how to extract different measures of the price disposition effect within our modelling framework and provide some numerical results. It is found that the inclusion of probability weighting can produce much more reasonable levels of the disposition effect under a range of asset performance, and this serves as an improvement over the existing asset sale models.

Chapter 2 considers the impact of allowing PT agents to follow randomised strategies in an exit-timing decision task. We demonstrate a feature which has not been considered to date. In presence of probability weighting, an agent may benefit from adopting a randomised strategy. In a finite horizon discrete model of casino gambling of Barberis (2012), we show that allowing an agent to follow randomised strategies can lead to strict improvement in the game value. Allowing randomised strategies also leads to drastic change in prediction in an infinite horizon continuous time setup. In an optimal stopping model under a general PT framework, Ebert and Strack (2015) show that a naive agent can always find a non-trivial gambling strategy at every wealth level which is strictly preferred to stopping immediately. From this, it is inferred that a naive agent will never stop voluntarily, and this casts doubts over the applicability of PT in a dynamic context. When randomised strategies are allowed, however, we show that the optimal strategy of a naive agent may involve stopping with positive probability. Through detailed analysis of two stylised examples as well as numerical studies on a more general model, we show that voluntary cessation of gambling with naive agents is possible which is a more realistic prediction.

A multi-asset Merton's investment and consumption problem with transaction costs is studied in Chapter 3. In general it is difficult to make analytical progress towards a solution in such a problem, but we specialise to a case where transaction costs are zero except for sales and purchases of a single asset which we call the illiquid asset. Leveraging

the analysis of Hobson et al. (2016) for the model with a single risky asset only, we show that the underlying HJB equation can be transformed into a first order boundary value problem. The optimal strategy is to trade the illiquid asset only when the fraction of the total portfolio value invested in this asset falls outside a fixed interval. Important properties of the multi-asset problem (including when the problem is well-posed, ill-posed, or well-posed only for large transaction costs) can be inferred from the behaviours of a quadratic function of a single variable and another algebraic function. We also discuss some comparative static results and their financial interpretations.

Chapter 1

Probability weighting and price disposition effect in an asset liquidation model

"Take care to sell your horse before he dies. The art of life is passing losses on."

— Robert Frost, *The Ingenuities of Debt*

1.1 Introduction

Despite the attractiveness of the principles of expected utility theory (EUT), it has long been recognised that it fails to fully explain individuals' attitudes towards risk. One of the most prominent alternatives to EUT is prospect theory (PT), originally proposed by Kahneman and Tversky (1979) and extended later by Tversky and Kahneman (1992). PT features the following key ingredients. First, utilities or values are derived in terms of gains and losses relative to a reference point rather than the final wealth level. Second, the value function exhibits concavity in the domain of gains and convexity in the domain of losses, and is steeper for losses than for gains to capture a phenomenon known as loss aversion. Finally, the most distinctive feature of PT is that cumulative probabilities are re-weighted such that individuals overweight tail events.

In recent years, probability weighting has been successfully linked, both theoretically and empirically, to a wide range of financial phenomena.[1] Barberis and Huang (2008) show that, in a financial market where investors evaluate risk according to prospect theory, probability weighting leads to the prediction that the skewness in an asset's return distribution will be priced. This idea has been used to explain low average returns of IPO stocks (Green and Hwang (2012)), the apparent overpricing of out-of-the-money options and the variance premium (Polkovnichenko and Zhao (2013), Baele et al. (2014)), the lack of diversification in household portfolios (Polkovnichenko (2005)), why riskier firms grant more stock options to non-executive employees (Spalt (2013)), and many other puzzles. On an aggregate scale, De Giorgi and Legg (2012) show that probability weighting is useful in generating a large equity premium - and can do so independently of loss aversion (Benartzi and Thaler (1995)). Probability weighting has also been helpful in understanding betting behaviour - the favourite long-shot bias (see Schneider and Spalt (2016) who show CEOs allocate capital with a long-shot bias) and the popularity of casino gambling (Barberis (2012)). In this chapter, we contribute to this broad agenda by showing that in the setting of dynamic models of investor trading, probability weighting can, in combination with the other elements of prospect theory, generate more realistic trading behaviour and satisfactorily explain an empirical anomaly known as the price disposition effect.

The disposition effect is one of the most robust effects in the empirical literature on investors' behaviours. It refers to the stylised fact that investors have a higher propensity to sell risky assets with capital gains compared to risky assets with capital losses (Shefrin and Statman (1985), Odean (1998), Genesove and Mayer (2001), Grinblatt and Keloharju (2001), Feng and Seasholes (2005), Dhar and Zhu (2006), Kaustia (2010), Jin and Scherbina (2011), Ben-David and Hirshleifer (2012), Birru (2015)). Odean's well known study computes the frequency with which individual investors sell winners and losers relative to opportunities to sell each and finds gains are realised at a rate around 50% higher than losses. Although prospect theory provides a leading explanation of the disposition effect (Odean (1998), Shefrin and Statman (1985)),[2] the literature linking the two is largely silent on the impact of the probability weighting feature of prospect theory. We fill this gap in the current chapter.

[1] See Barberis (2013) for a discussion and overview.

[2] Odean (1998) explicitly considers expected utility explanations for the asymmetry across winners and losers based on richer specifications of the investor's problem, finding that portfolio rebalancing, transaction costs, taxes, and rationally anticipated mean reversion cannot explain the observed asymmetry. Weber and Camerer (1998) find that incorrect beliefs concerning mean reversion cannot explain the disposition effect either.

The well known intuition from Shefrin and Statman (1985) linking PT to the disposition effect argues that PT's risk seeking over losses encourages investors to continue gambling when losing, whilst risk aversion over gains means investors tend to sell assets which have increased in value. Since this is a static argument, there has been a recent program in the literature attempting to build rigorous models formalising this link in a dynamic setting. Despite the intuition, it is a challenge for existing prospect theory models to explain the disposition effect (Kyle et al. (2006), Kaustia (2010), Barberis and Xiong (2009), Barberis and Xiong (2012), Henderson (2012), Li and Yang (2013)).[3] Indeed, the jury is still out. The difficulty is that although the convexity over losses and loss aversion do indeed act to encourage the investor to continue gambling in the domain of losses, this effect tends to be too strong. In many models the investor rarely (or even never) stops voluntarily, giving an extreme disposition effect.[4] Progress has been made by Ingersoll and Jin (2013) who study a realisation utility model with reference dependent S shaped preferences and show that consideration of reinvestment improves the range of parameters over which losses are taken. The model of Ingersoll and Jin (2013) gives an improved fit to the disposition effect, but requires considerable adjustments on the Tversky-Kahneman (TK) value functions and how they are applied.[5]

In this chapter we show the inclusion of probability weighting makes it easier for prospect theory to deliver a realistic level of the disposition effect. Intuitively, overweighting of extremely poor outcomes encourages the investor to stop-loss earlier in the loss region, while overweighting of extremely good outcomes provides him the incentive to let the profit run when winning. In isolation, therefore, probability weighting would work in the *opposite* direction to the disposition effect. When used *in tandem* with the other ingredients of PT (S shaped value function, loss aversion), probability weighting moderates the level of the disposition effect predicted by the model to give values which are much closer to observed

[3] There are other theories of the reference point that can potentially generate a disposition effect, such as a reference point given by a weighted average of recent prices (Weber and Camerer (1998), Odean (1998)), or by investors' expectations (Kőszegi and Rabin (2006), Meng and Weng (2016), Magnani (2015)).

[4] Most of this literature finds the investor never sells at a loss. An exception is Henderson (2012), who shows that under the Tversky and Kahneman (1992) value function, there is a loss threshold at which the investor will sell, but this only occurs for ranges of parameters where the stock has very poor expected returns, ie. where the investor gives up despite her loss aversion and convex preferences. For higher expected returns, an extreme disposition effect still emerges as loss aversion and convexity are dominant forces.

[5] First, the value function is applied over rates of return rather than dollar changes. Second, the TK value function is altered so that the marginal utility at the origin is finite. Further, an implausibly high risk seeking parameter is needed to obtain a good fit. To obtain a better fit for plausible parameters Ingersoll and Jin (2013) mix 50-50 realisation utility investors with random Poisson traders.

empirical levels. Indeed, we show the model can match Odean's measure of the disposition effect with realistic parameters.

Researchers studying the disposition effect have also recently examined how the rate of sale of stocks depends upon the relative magnitude of gains or losses (Feng and Seasholes (2005), Seru et al. (2010), Ben-David and Hirshleifer (2012), Barber and Odean (2013) and An (2016)). There is broad agreement amongst researchers that the estimated hazard rate, as a function of returns since purchase, is higher on gains than on losses. This is an evidence in favour of a disposition effect amongst investors because the higher level over gains means that the average propensity to sell is higher for gains than for losses. In our model we derive a trading rule - expressed as a price-level dependent selling rate per unit time - which is consistent with the optimal behaviour of a PT investor. We generalise this model-based selling intensity to multiple heterogeneous investors with different preferences. We demonstrate that with heterogeneous investors with TK value and weighting functions, and differing loss aversion levels, the model's implied selling rate matches the qualitative features of the empirical data including the disposition effect. Furthermore, the model is able to get reasonably close on magnitudes - including the daily probabilities of sale in the empirical data.

Probability weighting has two main impacts on the optimal trading strategy of a PT investor. First, the PT investor no longer aims for a simple threshold strategy on gains. Instead, probability weighting on gains encourages the investor to aim for a long-tailed distribution, placing some probability mass on extremely high gains precisely because these are the outcomes that are overweighted by the investor under probability weighting. Second, overweighting of extreme losses encourages the investor to stop - and thus we see a finite stop-loss threshold for a wider range of parameters than those found in the absence of probability weighting. Taken together, we show that the optimal target distribution of asset sale price for a PT investor consists of a single stop-loss threshold and a continuous distribution over gains. The distribution over gains is long right tailed, because the investor wishes to gamble on the very best returns, which he overweights. This can be contrasted to investor behaviour in PT models without probability weighting (Ingersoll and Jin (2013), Henderson (2012) and Barberis and Xiong (2012)) which produce two-sided thresholds. The new solution structure could explain two other well documented phenomena: the use of trading strategies which are of stop-loss form but not stop-gain and the gambling preferences implicit in the investment choices of retail investors for right-skewed asset returns.

Underpinning our results on the disposition effect is the important technical progress we make on the form of the optimal prospect for a PT investor. We work in a continuous time, infinite horizon, dynamic optimal stopping model of asset trading where investors have PT preferences and probability weighting. A mathematical building block for our work is Xu and Zhou (2013) who also focus on characterising the optimal strategy for a PT investor who can precommit but most of their results are for gains (or losses) separately.[6] Our results are proved under a condition on the value function and the weighting function applied to losses which we call the elasticity condition. Under this assumption, and with general asset price processes, we characterise the investor's optimal prospect or distribution as a single loss threshold together with a continuous distribution over gains.[7] This characterisation is valid for all of the popular functions in the literature, including the value and weighting functions of Tversky and Kahneman (1992) since each of these specifications satisfies the elasticity hypothesis.

This chapter is structured as follows. In Section 1.2 we formulate our problem of asset liquidation under PT preferences and state the main result. We take the Tversky and Kahneman value and weighting functions as the base model in Section 1.3 and provide some numerical results to highlight the features of the optimal trading strategies. To measure the disposition effect within our theoretical model, we present two approaches in Section 1.4 and Section 1.5 respectively based on the Odean's disposition ratio and a selling intensity function. Finally, we give our closing remarks in Section 1.6. Proofs and some supplementary results are given in an appendix.

1.2 Model of asset liquidation under prospect theory preferences

1.2.1 Elements of prospect theory

Under prospect theory, utility is evaluated in terms of gains and losses relative to a reference point, rather than over final wealth. Denote by Z a random variable and by R the reference

[6]Although Xu and Zhou (2013) consider a variety of shapes of value and weighting functions, they do not treat Tversky and Kahneman (1992) inverse-S shaped weighting functions together with S shaped utility. Furthermore, although we can deduce from their results that the optimal prospect on losses places mass on at most three points, they do not obtain a single loss threshold, and they do not consider implications for the disposition effect. Thus, they cannot speak to the questions we answer in this chapter.

[7]We believe this structure for the solution holds more generally, but the elasticity condition provides a simple sufficient condition which can be checked on the value function and probability weighting function separately. This decoupling makes the sufficient condition relatively simple to check.

point or level and let $Y = Z - R$ denote the gain or loss relative to the reference level. Let U be the (continuous, strictly increasing, twice differentiable away from zero) utility or value function defined over the range of Y such that $U(0) = 0$. Under prospect theory, U is concave over gains and convex over losses. It also exhibits loss aversion, whereby a loss has a larger impact than a gain of equal magnitude.

Kahneman and Tversky (1979) propose power functions of the form

$$U(y) = \begin{cases} y^{\alpha_+}, & y \geqslant 0, \\ -k(-y)^{\alpha_-}, & y < 0, \end{cases} \tag{1.1}$$

where $0 < \alpha_\pm < 1$. The parameters $1 - \alpha_+$ and $1 - \alpha_-$ represent the coefficients of risk aversion and risk seeking, respectively. The parameter $k \geqslant 1$ governs loss aversion, introducing an asymmetry about the origin. Experimental results of Tversky and Kahneman (1992) give estimates of $\alpha_+ = \alpha_- = 0.88$ and $k = 2.25$.

The final ingredient of prospect theory is that the probabilities of extreme events are overweighted where the degree of distortion can differ for gain and loss outcomes. Let $w_\pm : [0,1] \mapsto [0,1]$ be a pair of (continuous, strictly increasing, differentiable) probability weighting functions with $w_\pm(0) = 0, w_\pm(1) = 1$. Then the prospect theory value of Z is given by (see Kothiyal et al. (2011))

$$\mathcal{E}(Z) = \int_0^\infty w_+(\mathbb{P}(U(Z-R) > y))dy - \int_{-\infty}^0 w_-(\mathbb{P}(U(Z-R) < y))dy. \tag{1.2}$$

Many experimental studies (e.g. Camerer and Ho (1994), Wu and Gonzalez (1996), Tversky and Kahneman (1992)) and recent empirical estimates (Polkovnichenko and Zhao (2013)) find that individuals typically overweight the events in the tails of the distribution. Overweighting of small probabilities on extreme events suggests the probability weighting functions $w_\pm$ should be inverse-S shaped functions. In particular there exist $q_\pm$ such that $w_\pm$ is concave on $[0, q_\pm]$ and convex on $[q_\pm, 1]$. Moreover, experimental studies typically demonstrate that $w_\pm(1/2) < 1/2$. Tversky and Kahneman (1992) propose the probability weighting functions

$$w_\pm(p) = \frac{p^{\delta_\pm}}{(p^{\delta_\pm} + (1-p)^{\delta_\pm})^{1/\delta_\pm}} \tag{1.3}$$

for $0.28 < \delta_\pm \leqslant 1$.[8] Median estimates of $\delta_\pm$ are reported to be $\delta_+ = 0.61$ and $\delta_- = 0.69$. Alternative forms of $w_\pm$ proposed in the literature include Goldstein and Einhorn (1987) and Prelec (1998).[9]

[8] A lower bound on $\delta_\pm$ is required to ensure monotonicity of $w_\pm$.

[9] The Goldstein and Einhorn (1987) and Prelec (1998) weighting functions are given by:

$$w_\pm^{GE}(p) = \frac{\gamma_\pm p^{d_\pm}}{\gamma_\pm p^{d_\pm} + (1-p)^{d_\pm}}, \qquad w_\pm^{P}(p) = \exp(-b_\pm(-\ln p)^{a_\pm}) \tag{1.4}$$

1.2.2 Model formulation

Our model is a partial equilibrium framework with an infinite horizon. An investor holds an asset whose price at time t is given by P_t. He can sell or liquidate the asset at any time in the future. At the liquidation time τ of his choice,[10] he receives the price P_τ and compares it to his reference level R, which may be the breakeven level or price paid for the asset.[11] The realised gain or loss to the investor at the sale time is the difference between the asset price and reference level, $P_\tau - R$, which he evaluates at the outset by (1.2) on setting $Z = P_\tau$.

The goal of the investor is to choose the best time τ to sell the asset to maximise the PT value:[12]

$$\sup_{\tau} \left(\int_0^{\infty} w_+ \left(\mathbb{P} \left(U \left(P_\tau - R \right) > y \right) \right) dy - \int_{-\infty}^{0} w_- \left(\mathbb{P} \left(U \left(P_\tau - R \right) < y \right) \right) dy \right). \tag{1.5}$$

Note that if $w_\pm(p) = p$ so that there is no probability weighting, we recover the model of Henderson (2012) (see also Kyle et al. (2006)).

In general, we can model the asset price $P = (P_t)_{t \geqslant 0}$ by a time-homogeneous diffusion with state space $\mathcal{J}$, given by

$$dP_t = \kappa_P(P_t)dt + \sigma_P(P_t)dB_t. \tag{1.6}$$

Here $B = (B_t)_{t \geqslant 0}$ is a standard Brownian motion and $\kappa_P : \mathcal{J} \to \mathbb{R}$ and $\sigma_P : \mathcal{J} \to (0, \infty)$ are Borel functions. We assume $\mathcal{J}$ is an interval with endpoints $-\infty \leq a_J < b_J \leq \infty$ and that P is regular in (a_J, b_J). We will later specialise to the most popular asset price specification where P is a geometric Brownian motion (or equivalently P is lognormal). In that case $\mathcal{J} = (0, \infty)$, $\kappa_P(p) = \kappa p$ and $\sigma_P(p) = \sigma p$ for constants κ and σ.

Following Henderson (2012) it is convenient to reformulate the objective (1.5) by transforming the asset price into a martingale. We define $X_t := s(P_t)$ where the scale function s ensures X is a (local) martingale.[13] We are free to normalise s such that $s(R) = 0$,

respectively, for parameters $0 < \gamma_\pm < 1$, $0 < d_\pm < 1$ and $a_\pm > 0$, $0 < b_\pm \leq 1$.

[10] The liquidation time τ must be a stopping time.

[11] In common with the prospect theory models we compare to, we consider a fixed reference level.

[12] In common with Kyle et al. (2006), Henderson (2012), we take zero interest rate for tractability reasons. If we were to include time discounting, the discounting of losses will encourage sale delay, but this will not be due to the prospect theory preferences or probability weighting. Indeed, Barberis and Xiong (2012) show a piecewise linear realisation utility investor with positive time discounting will never sell at a loss voluntarily.

[13] The scale function of P can be identified as the increasing, non-degenerate solution (which is unique up to positive affine transformation) to the ordinary differential equation

$$\frac{1}{2}\sigma_P^2(p)s''(p) + \kappa_P(p)s'(p) = 0.$$

Then $X = (X_t)_{t \geq 0}$ defined by $X_t := s(P_t)$ is a (local) martingale. We assume that $\kappa_P(.)$ and $\sigma_P(.)$ are and hence $X_t \geqslant 0$ when $P_t \geqslant R$ and $X_t \leqslant 0$ when $P_t \leqslant R$. Then X_t represents the transformed gains and losses relative to the reference level. If we take the reference level to be the initial asset price $R = P_0$, then $X_0 = 0$.

The state space of X is an interval with endpoints $L = s(a_J)$ and $M = s(b_J)$. Then $L < 0$ represents the potential maximum loss. We assume $L > -\infty$ to ensure the problem is non-degenerate or well-posed. Define

$$v(x) := U\left(s^{-1}(x) - R\right) = U\left(p - R\right).$$

Then the investor's objective (1.5) can be rewritten as

$$\sup_{\tau}\left(\int_0^{v(M)=U(b_J-R)} w_+\left(\mathbb{P}\left(v(X_\tau) > y\right)\right) dy - \int_{U(a_J-R)=v(L)}^{0} w_-\left(\mathbb{P}\left(v(X_\tau) < y\right)\right) dy\right). \tag{1.7}$$

One of the insights of Xu and Zhou (2013) is that the argument in (1.7) only depends on the law of X_τ. Hence, (1.7) can in turn be rewritten as

$$\sup_{\nu\in\mathcal{A}}\left(\int_0^{v(M)} w_+\left(1 - F_\nu(v^{-1}(y))\right) dy - \int_{v(L)}^{0} w_-\left(F_\nu(v^{-1}(y))\right) dy\right) \tag{1.8}$$

where $\mathcal{A}$ is the set of attainable laws of X_τ and F_ν is the cumulative distribution function of the law ν. The set of attainable laws $\mathcal{A}$ can be characterised by $\mathcal{A} = \{\nu : \int y\nu(dy) = X_0\}$.[14] Our investor evaluates (1.8) at the outset and commits today to achieve the desired target distribution or prospect.

1.2.3 Tversky and Kahneman (1992) value and weighting functions: the base model

Our aim, once we have some general characterisations to guide us, is to apply our results to study the most popular value and weighting functions. We will consider the Kahneman and Tversky (1979) value function in (1.1), which is of piecewise power S shape together

sufficiently regular that there exists a weak solution to the stochastic differential equation (1.6) and that the scale function s exists (see Revuz and Yor (1999)).

[14] At this point the fact that X is a local martingale is important since it allows us to give a simple characterisation of the space of attainable laws. Since $L > -\infty$ such that X is bounded below, it is a supermartingale and any attainable law ν must satisfy $\int y\nu(dy) = \mathbb{E}[X_\tau] \leq X_0 = s(P_0)$; conversely the theory of Skorokhod embeddings tells us that for every law ν with $\int y\nu(dy) \leq X_0$ there is a stopping rule τ such that $X_\tau \sim \nu$. Finally, since U is increasing, in searching for the supremum in (1.8) we may restrict attention to laws satisfying $\int y\nu(dy) = X_0$. Hence we may set $\mathcal{A} = \{\nu : \int y\nu(dy) = X_0\}$. If ν^* is the optimal law arising in (1.8), so that the optimal prospect for the process in natural scale is ν, then the optimal prospect for P has law μ^* where $F_{\mu^*}(p) = F_{\nu^*}(s(p))$.

with the Tversky and Kahneman (1992) inverse-S shaped weighting functions given in (1.3). In the base model we will take the price to follow a geometric Brownian motion, so that $P = (P_t)_{t \geq 0}$ solves

$$dP_t = P_t(\kappa dt + \sigma dB_t) \tag{1.9}$$

for constant expected return κ and volatility σ with $\kappa < \sigma^2/2$. The hypothesis that $\kappa < \sigma^2/2$ ensures the price does not reach arbitrarily high levels with probability one. (The case $\kappa \geq \sigma^2/2$ leads to a degenerate problem whereby the investor never sells.)

Define the constant parameter $\beta := 1 - \frac{2\kappa}{\sigma^2}$ which involves the return-for-risk-per-unit-variance κ/σ^2 and thus reflects the expected performance of the asset. We assume $\kappa \geq 0$ so that in expectation P is non-decreasing and then $\beta \leq 1$. Our assumption $\kappa < \sigma^2/2$ implies that $\beta > 0$. The scale function is given by $s(p) = p^\beta - R^\beta$. Then, since $\beta > 0$ we have $L = s(a_J) = s(0) = -R^\beta > -\infty$, consistent with our non-degeneracy assumption, and the scaled value function is given by

$$v(x) = U(s^{-1}(x) - R) = \begin{cases} \left((x + R^\beta)^{1/\beta} - R\right)^{\alpha_+}, & x \geqslant 0, \\ -k\left(R - (x + R^\beta)^{1/\beta}\right)^{\alpha_-}, & -R^\beta \leq x < 0, \end{cases} \tag{1.10}$$

where we use $M = s(\infty) = \infty$.

In addition to $\beta > 0$ we require some further restrictions on parameter values to avoid situations leading to infinite PT value and to obtain a well defined optimal strategy. First, if the growth rate or Sharpe ratio on the asset is too large relative to risk aversion, the investor simply waits indefinitely to take advantage of the favourable asset. To rule this out, we need that $\alpha_+ < \beta$.[15] Under this assumption (together with the condition that $\beta \leqslant 1$) it can be checked by differentiation that v is concave on $[0, \infty)$ and convex on $[L, 0]$. In fact, the following result shows that for well-posedness we require a slightly stronger condition incorporating the strength of probability weighting on gains. The proof is given in Appendix 1.C.

Proposition 1.1. *In our base case model with Tversky and Kahneman (1992) value and weighting functions, $\alpha_+/\delta_+ < \beta$ is a sufficient condition for there to be a finite value function and a well defined optimum. On the other hand, if $\alpha_+/\delta_+ > \beta$ then the problem is ill-posed.*

The inclusion of probability weighting over gains has increased the set of scenarios whereby the investor waits indefinitely. Rewriting the condition for a non-degenerate solu-

[15] A similar condition also arises in standard infinite horizon portfolio problems.

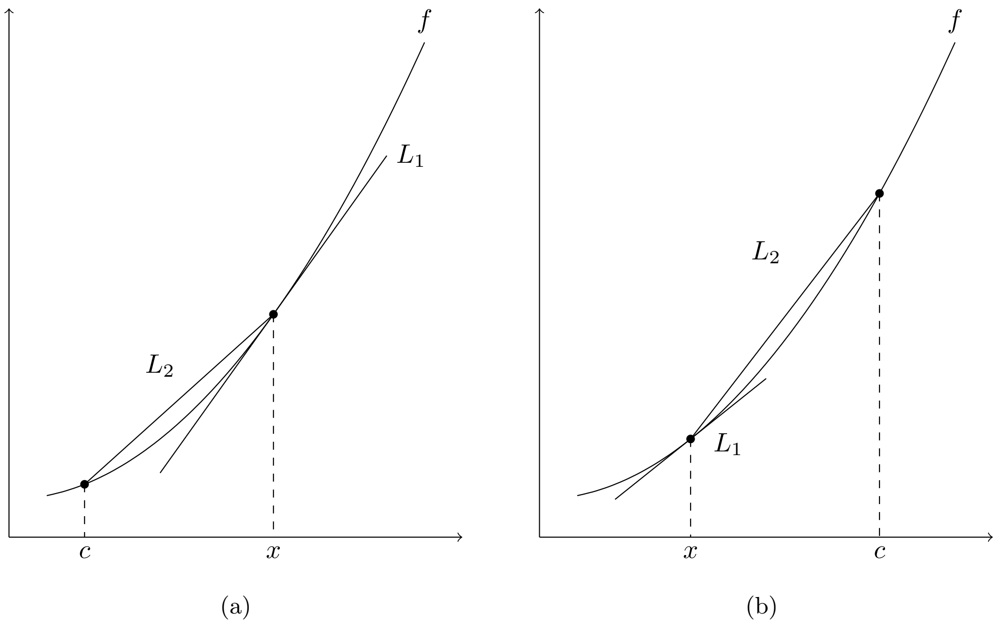


Figure 1.1: A graphical illustration of the elasticity measure $E(x; f, c)$. The left and right sketch correspond to cases with $x > c$ and $x < c$ respectively. Consider a reference point on f given by $(c, f(c))$. Then $\frac{f(x)-f(c)}{x-c}$ is the slope of the straight line L_2 joining the reference point and $(x, f(x))$, while $f'(x)$ is the slope of the tangent to f at x denoted by L_1. Hence, for a fixed c, $E(x; f, c)$ can be interpreted as the ratio of slope of L_1 and L_2 as x varies.

tion as $\delta_+ > \alpha_+/\beta$, we see that we cannot have probability weighting over gains to be too strong, as this will cause the investor to simply continue waiting.

1.2.4 Elasticity measure

We end this section by introducing an elasticity measure which will be useful in allowing us to characterise the optimal solution.

Definition 1.2. *For a monotonic and continuously differentiable function $f : S \to \mathbb{R}$, the elasticity measure (parameterised by x) relative to a reference point c is defined as*

$$E_{f,c}(x) = E(x; f, c) = \frac{(x-c)f'(x)}{f(x)-f(c)} = \frac{f'(x)}{\frac{f(x)-f(c)}{x-c}}$$

where $x, c \in S$ and $x \neq c$. At $x = c$, and provided $f'(c) \neq 0$, we define $E(c; f, c) = 1$ by L'Hôpital's rule.

For a graphical interpretation of this measure, see Figure 1.1.

The elasticity measure in Definition 1.2 has several useful properties, the derivations of which are straightforward, but are given in Appendix 1.D.

Proposition 1.3. *Let ι denote the identity function $\iota(x) = x$, and suppose f and g are monotonic and continuously differentiable. Let $a \neq 0$ and b be constants. Then $E(x; a\iota + b, c) = 1$ and $E(x; g \circ f, c) = E(x; f, c)E(f(x); g, f(c))$.*

The above results are key in proving the following.

Proposition 1.4.

1. *Suppose $0 < \alpha_- < 1$ and $0 < \beta \leq 1$. If $v(x) = -k\left(R - (x + R^\beta)^{1/\beta}\right)^{\alpha_-}$ for $-R^\beta = L \leq x \leq 0$ then $E(x; v, c)$ is increasing in x for $x \in [L, 0]$ for fixed $c \in [L, 0)$.*
2. *If the weighting function w is of the form proposed by Tversky and Kahneman (1992), Goldstein and Einhorn (1987) or Prelec (1998) and has inflexion point q then $E(p; w, r)$ is decreasing in p for $0 \leq p \leq \min\{r, q\}$ for any r in $[0, 1]$.*

1.2.5 The main result

In what follows we assume that U is S shaped, $w_\pm$ is inverse-S shaped with inflexion point $q_\pm$, $L = s(a_J) > -\infty$ and $M = s(b_J) = \infty$. Recall the definition $v(x) := U(s^{-1}(x) - R)$.

Assumption 1.5 (S shaped assumption on v)**.** *v is concave on $[0, \infty)$ and convex on $[L, 0]$. Further, $v'(0+) = \infty$ and $\lim_{x\uparrow\infty} v'(x) = 0$.*

Note that this assumption is satisfied in our base case, and more generally whenever $U'(0+) = \infty$, $\lim_{p\uparrow b_J=\infty} U'(p) = 0$ and P is a martingale whence the scale function is the identity function. More generally it depends on the interplay between the value function U and the dynamics of the price process.

Assumption 1.6 (Elasticity assumption)**.** *$E(x; v, L)$ is increasing in x for $x \in [L, 0]$ and $E(p; w_-, r)$ is decreasing in p for $0 \leq p \leq \min\{r, q_-\}$ for any r in $[0, 1]$.*

By Proposition 1.4 both parts of the Elasticity Assumption are satisfied in the base case model. They are satisfied for a range of other probability weighting and value functions as well.

Our main theoretical result is the following where a full proof is presented in Appendix 1.B.

Proposition 1.7. *Suppose Assumptions 1.5 and 1.6 hold. Then the optimal prospect has a distribution which consists of a point mass in the loss regime and a point mass at some* *point a in the gains regime, together with a continuous distribution on the unbounded interval (a, ∞).*

More precisely, the quantile function of the optimal prospect P_τ is given by $G_P := s^{-1} \circ G_X$ where G_X represents the quantile function of the optimal scaled prospect $X_\tau = s(P_\tau)$ taking the form

$$G_X(u) = \begin{cases} -\frac{1}{1-\phi}\left[\int_0^\phi (v')^{-1}\left(\frac{\lambda}{w'_+(\psi \wedge y)}\right) dy - X_0\right], & u \leq 1-\phi, \\ (v')^{-1}\left(\frac{\lambda}{w'_+(\psi)}\right), & 1-\phi < u \leq 1-\psi, \\ (v')^{-1}\left(\frac{\lambda}{w'_+(1-u)}\right), & 1-\psi < u \leq 1, \end{cases} \tag{1.11}$$

for some $\lambda > 0$, $\phi \in [0,1]$, $\psi \leq q_+ \wedge \phi$ such that

$$X_0^+ \leq \int_0^\phi (v')^{-1}\left(\frac{\lambda}{w'_+(\psi \wedge y)}\right) dy \leq X_0 - (1-\phi)L.$$

It follows that the optimal strategy for a PT investor is a stop-loss combined with a strategy yielding a long-tailed distribution on gains.

1.3 Numerical results under the base model

We now examine the optimal sales strategies for the investor with Tversky and Kahneman value and weighting functions. Tversky and Kahneman (1992) estimated the preference parameters as: $\alpha_+ = \alpha_- = 0.88$, loss aversion $k = 2.25$. The TK parameters arise from experimental settings with small gamble sizes and we would expect higher levels of risk aversion in a financial trading setting. Wu and Gonzalez (1996) estimate $\alpha_+ = 0.5$ when they use the TK parameterisation. Furthermore, Ingersoll and Jin (2013) consider $\alpha_+ = 0.5, \alpha_- = 0.9$ as one of their base parameter sets. For consistency, we will also adopt $\alpha_+ = 0.5$, $\alpha_- = 0.9$ as our base case. Our base loss aversion parameter level is $k = 1.25$. For all parameters we will consider a range of values when we look at comparative statics. Estimates of the TK probability weighting parameters have been quite consistent across experimental and empirical studies. TK estimate the probability weighting parameters as $\delta_+ = 0.61, \delta_- = 0.69$. Wu and Gonzalez (1996) find experimentally that $\delta_+ = 0.71$. Recently Baele et al. (2014) estimate the degree of probability weighting from S&P 500 equity and option data and report a range of 0.72-0.79. Reflecting these findings, we take base parameters of $\delta_+ = \delta_- = 0.7$.

As described in Proposition 1.7, the optimal prospect consists of a single loss threshold together with a distribution over gains. Figure 1.2 illustrates the results for our base set of parameters. Note the reference level is taken to be $R = P_0 = 1$, so prices above 1 represent gains and below 1, losses. In panel (a), we display the optimal quantile function. The distribution on losses is a point mass. The investor places just over 0.4 of the probability mass onto the single loss threshold of about 0.7, which is a 30% loss relative to the reference level. The remainder of the probability mass is distributed over the gains, starting just above the reference level of one, and tailing off at around 2.4. (More precisely, the level 2.4 represents the upper 99th percentile of the distribution.) We see the distribution over gains is highly positively skewed, in that the investor puts most weight on the value of the stock close to the reference level, and the distribution has a long right tail. We will take a closer look at skewness measures in Section 1.3.2.

We first recap the form of the solution in the absence of probability weighting, when $\delta_+ = \delta_- = 1$. In this case (see Henderson (2012)), the optimal strategy is a threshold sale strategy. There will be a gain threshold level and a loss threshold level, and the optimal strategy is to stop the first time the price process leaves this interval. The corresponding prospect is a distribution on exactly two points. Typically the gain threshold is very close to the reference level. For some price parameters, the loss threshold is at zero, and it is never optimal to sell at a loss. (Instead, sale is postponed indefinitely). For other parameters, there will be a loss threshold, which is usually much further from the reference level than the gain threshold. Thus, if losses are realised, they are much larger in size than gains. Why? The marginal utility of a gain or loss is decreasing with size, so small gains and large losses are preferable. When the asset price parameters are such that $\beta \leq 1$, there is no lower loss threshold and the investor avoids voluntary losses. This is the case in panel (b) of Figure 1.2 for $\delta_+ = \delta_- = 1$. In this case, the convexity of the utility and loss aversion together mean that the investor prefers to continue to gamble and delay any losses.

Now we can look at the impact of probability weighting on the distribution. In panel (b), the probability weighting parameter $\delta_+ = \delta_-$ is varied on the x-axis. For varying values of $\delta_+ = \delta_-$, we display the single loss threshold together with the lower bound and upper 99th percentile of the distribution over the gains regime. Note that a loss threshold of zero in the figure effectively represents the situation where the investor never voluntarily realises losses. The vertical dashed line on the figure represents the base parameters for probability weighting and thus corresponds to the values used in panel (a).

We can now see the key impact of probability weighting. As we introduce probability weighting by reducing the values of δ_+, δ_-, we see the optimal prospect on gains completely changes character and switches from a point mass to a distribution with unbounded support. The tail of this distribution gets larger as probability weighting increases in strength. Why is this the case? Risk aversion alone makes small gains attractive. However now the investor overweights extreme events - in particular extreme gains - and this encourages him to place some probability mass on these extreme wins. His distribution over gains is right-skewed in that most mass is still concentrated on lower gain levels, but probability weighting causes him to want to gamble on the best wins by placing some mass there. As probability weighting becomes stronger, the investor places mass on more and more extreme wins and hence the upper 99th percentile increases as δ_+ decreases.

As probability weighting becomes sufficiently large ($\delta_\pm$ below about 0.75 in panel (b) thus including our base parameter of 0.7), we see that there will also be a strictly positive lower loss threshold at which the investor voluntarily takes losses. There are two forces driving this. First, the convexity and loss aversion are encouraging the investor to wait and avoid taking a loss. But now the investor overweights extreme events - in particular - extreme losses - which encourages him to cut-losses at some threshold. Importantly, the parameter region where a non-trivial loss threshold is present includes the levels of probability weighting commonly estimated in experimental and empirical studies.

We now return to the optimal distribution generated by our model and perform some comparative statics. We focus on the analogs of Panel (b) in Figure 1.2, and the location of the threshold on losses, together with the location of point mass on gains, and the 99% upper quantile of the distribution. In Figure 1.3 we investigate the impact of the probability weighting on gains and losses separately, the effect of risk aversion on gains and risk seeking on losses, and the impact of loss aversion on the investor's behaviour. Panels (a)-(e) plot the investor's optimal distribution as we vary each of the parameters in turn. In each panel, we indicate the location of the base parameter with a vertical dashed line.

Panels (a) and (b) vary one of the probability weighting parameters whilst keeping the other fixed. We see that it is δ_- that is governing whether there is a lower loss threshold and that we need enough weighting but not too much. If probability weighing on losses is not sufficiently strong, then the investor never takes a loss. However, panel (b) also shows that if δ_- is too low, ie. weighting on losses is too strong, then again, the investor never stops at a loss threshold.

Panels (c) and (d) vary the risk aversion and seeking parameters separately, whilst holding all other parameters fixed at their base values. In panel (c) we observe that higher levels of risk aversion results in the distribution over gains being pulled down closer to the reference level. If risk aversion over gains is sufficiently strong, below about 0.4 in the panel, the investor no longer realises losses. At the other extreme, we know that if risk aversion over gains is not strong enough, it violates the condition in Proposition 1 and the investor instead waits indefinitely. For the parameters in the graph, this would occur for values of $\alpha_+ > 0.63$. In particular, we see that under the original TK parameters $\alpha_+ = \alpha_- = 0.88$ (particularly low levels of risk aversion/seeking), the investor violates the condition in Proposition 1.1 and thus waits forever to sell, unless the expected return on the asset is unrealistically large and negative. A similar observation has been made by Ingersoll and Jin (2013) in the absence of probability weighting.

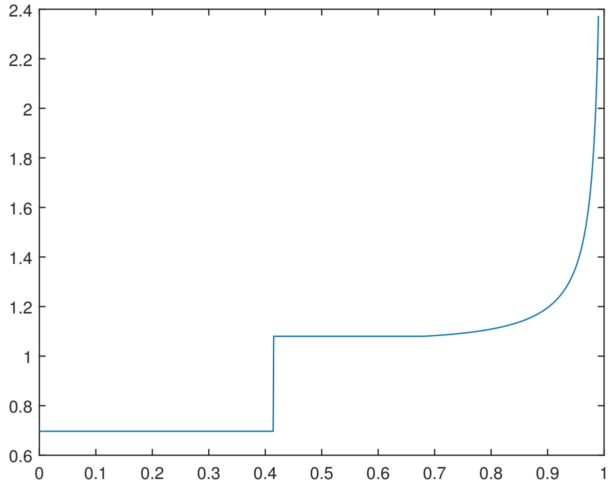
In panel (d) we see that as we increase the risk seeking parameter on losses (decrease α_-) the convexity of the utility becomes stronger and encourages the investor to avoid taking a loss. Precisely where the convexity becomes the dominant force will depend upon other parameters. If the asset was less attractive, then the investor would sell at a loss threshold for a larger range of α_-, and if the probability weighting on losses was stronger (but not too strong), the investor would again stop at a loss threshold for a larger range of α_-.

In panel (e) we vary loss aversion. As loss aversion becomes stronger, the investor chooses a loss threshold which is closer to the reference level as he is less willing to wait in the domain of losses. Larger values of k also reduce the long right-tail on gains.

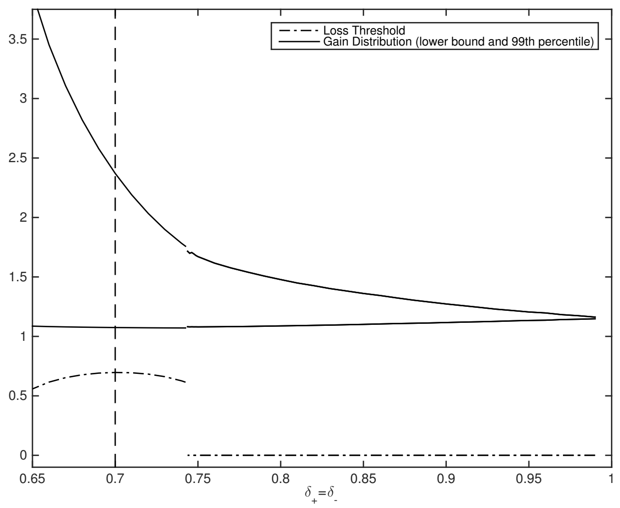
1.3.1 PT trading and stop-loss strategies

We have shown that a prospect theory investor with S shaped utility, loss aversion and probability weighting will trade to achieve a distribution over gains but will desire a stop-loss threshold over losses. This difference in how the investor trades gains and losses matches very well how investors behave in financial markets. Stop-loss strategies are in widespread usage in practice but stop-gain or take-gain strategies are much rarer.

Despite the popularity of stop-loss strategies in financial markets, they are not that easily justified by financial theory. Kaminski and Lo (2014) derive the impact of a stop-loss rule on the return characteristics of a portfolio and find stop-loss rules can increase the expected return if returns are non-random walks. Shefrin and Statman (1985) discuss stop-loss strategies in the context of self control - a stop-loss allows an investor to make loss realisation at a predetermined point automatic. Fischbacher et al. (2015) test this idea by investigating in a laboratory experiment whether the option of automatic selling devices causally reduces investors' disposition effect. Investors who had access to the automatic selling devices had significantly smaller disposition effects, which was driven by a significant increase in realised losses. They show it is the opportunity to ex ante commit to selling losses, which reduces the disposition effect. In contrast, neither the proportion of winners realised,

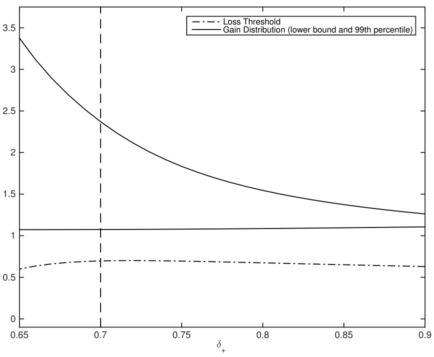


(a)

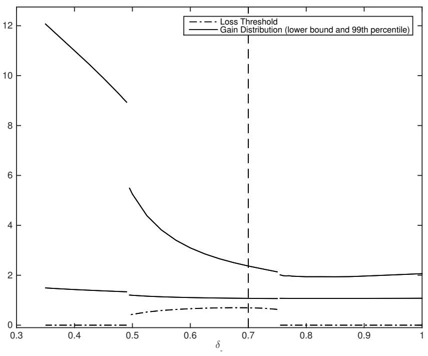


(b)

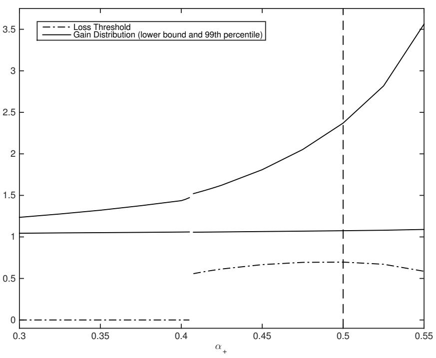
Figure 1.2: The optimal distribution in the model with Tversky and Kahneman value and weighting functions. The optimal distribution consists of a single loss threshold together with a distribution over the gains region. In panel (a), we display the optimal quantile function with asset price P on the y-axis. In panel (b), the probability weighting parameter $\delta_+ = \delta_-$ is varied on the x-axis. Displayed are the single loss threshold together with the lower bound and upper 99th percentile of the distribution over gains. The vertical dashed line indicates the base parameter value of $\delta_\pm = 0.7$ corresponding to panel (a). Base parameters used in both panels are $\alpha_+ = 0.5$, $\alpha_- = 0.9$, $\delta_+ = \delta_- = 0.7$, $k = 1.25$, $\beta = 0.9$, $R = 1$ and $P_0 = 1$.



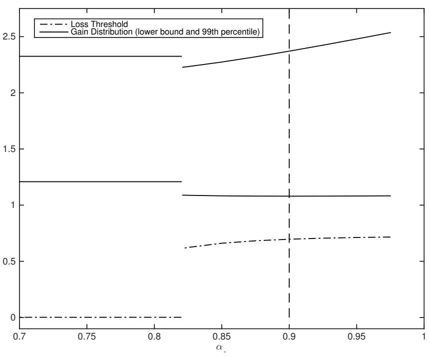
(a) δ_+ varying



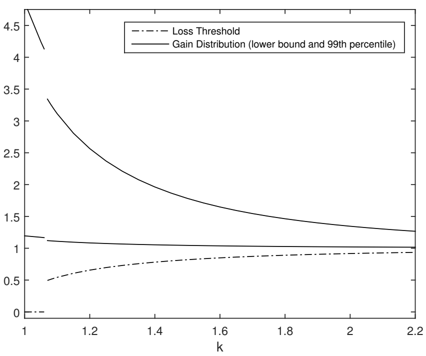
(b) δ_- varying



(c) α_+ varying



(d) α_- varying



(e) k varying

Figure 1.3: Comparative statics with respect to parameters with Tversky and Kahneman value and weighting functions. The optimal distribution consists of a single loss threshold together with a distribution over the gains region. In each panel, we display the single loss threshold together with the lower bound and upper 99th percentile of the distribution over the gains regime. Each panel varies one parameter at a time, keeping the others fixed at base values. The vertical line marks the location of the relevant base parameter in each panel. Base parameters used are $\alpha_+ = 0.5$, $\alpha_- = 0.9$, $\delta_+ = \delta_- = 0.7$, $k = 1.25$, $\beta = 0.9$, $R = 1$ and $P_0 = 1$.

nor the size of realised gains differed significantly across the treatments with automatic limits or no limits.

Standard expected utility settings can predict a stop-gain threshold at which an investor should sell but tend to put any lower threshold at $-\infty$ (see Viefers and Strack (2014)). Realisation utility (Barberis and Xiong (2012)) predicts a stop-gain threshold which resets each time a sale is made. If models do predict a stop-loss, they tend to also predict a stop-gain. This is the case in several prospect theory models without probability weighting - for instance, Henderson (2012) and Ingersoll and Jin (2013).

In contrast, we have shown a PT investor with probability weighting finds a stop-loss desirable but does not want to place a stop-gain. Instead, probability weighting encourages him to gamble on obtaining extreme (overweighted) gains. This fundamental difference in behaviour with regard to losses and gains in our model mirrors very well what we see in the financial markets.

1.3.2 PT trading and skewness

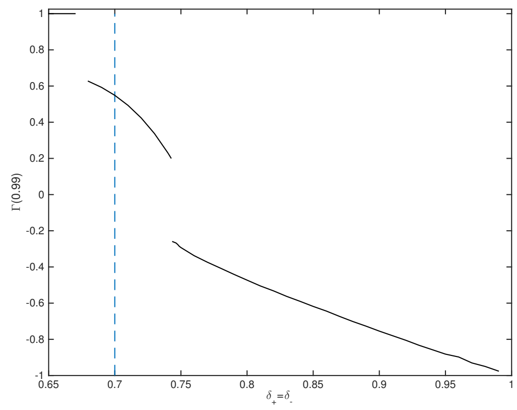
Prospect theory and skewness are heavily linked in the extant literature. Barberis and Huang (2008) show that in a financial market where investors evaluate risk according to prospect theory, probability weighting leads to the prediction that skewness in an asset's return distribution will be priced. Spalt (2013) argues using probability weighting that firms can use stock options to benefit from catering to an employee demand for lottery-like payoffs. Ebert and Hilpert (2015) show a strong preference for skewness contributes to the attractiveness of technical trading.

To demonstrate the role of probability weighting on skewness in our model, we calculate a measure of skewness for the optimal distribution under our base model with Tversky and Kahneman value and weighting functions. We use the robust, tail or quantile based measure of skewness of Hinkley (1975) (see Ebert and Hilpert (2015), Green and Hwang (2012) and Conrad et al. (2013))

$$\Gamma(0.99) = \frac{F^{-1}(0.99) + F^{-1}(0.01) - 2F^{-1}(1/2)}{F^{-1}(0.99) - F^{-1}(0.01)}$$

where F is the cumulative distribution function. Note that skewness is an attempt to summarise the shape of a distribution in a single statistic, which is often a difficult task. $\Gamma(0.99)$ depends only on the quantiles at 0.01, 0.5 and 0.99.

In Figure 1.4 we plot skewness, as measured by $\Gamma(0.99)$, across different levels of the probability weighting parameter $\delta_{\pm}$. The corresponding optimal distribution for the same



(a)

Figure 1.4: Skewness measure for the optimal distribution in the model with Tversky and Kahneman value and weighting functions. The skewness measure $\Gamma(0.99)$ is plotted for varying values of the probability weighting parameter $\delta_{\pm}$. The vertical dashed line indicates the base parameter value of $\delta_{\pm} = 0.7$. Other base parameters are $\alpha_{+} = 0.5$, $\alpha_{-} = 0.9$, $k = 1.25$, $\beta = 0.9$, $R = 1$ and $P_0 = 1$.

parameter choice was displayed earlier in Figure 1.2. We first observe that skewness can be positive or negative, and can take values over the full range of +1 and -1, depending on the level of probability weighting.

Without probability weighting, PT investors take small gains frequently, with some occasional large losses (Henderson (2012), Ingersoll and Jin (2013)). This typically leads to a left or negatively skewed distribution. In particular, $F^{-1}(0.01)$ is zero, whereas both $F^{-1}(0.5)$ and $F^{-1}(0.99)$ are equal and both just above P_0 (the agent follows a two-sided threshold strategy). It follows $\Gamma(0.99) = -1$ when $\delta_{\pm} = 1$. With an S shaped utility and no probability weighting, investors prefer left skewed return distributions.

Once probability weighting is included, the skewness measure is no longer -1. The investor does not follow a two-sided threshold and he looks for a long-tailed distribution on gains. This tail gets larger as $\delta_{\pm}$ decreases, although for $\delta_{\pm}$ close to one, the return distribution remains negatively skewed. For $\delta_{\pm}$ greater than about 0.75 the optimal prospect includes an atom at zero and the skewness statistic is negative. However, at $\delta_{\pm} \sim 0.75$ the optimal prospect undergoes a step change and the mass on losses moves from zero to a strictly positive level. This leads to a jump in the skewness statistic, which now becomes positive. As $\delta_{\pm}$ decreases further, the right tail on the optimal prospect becomes larger and the skewness increases further from about 0.2 to 0.6. Now the investor is taking losses of moderate size, regular small gains, and occasional large gains.

The second jump in the skewness statistic occurs when the total mass on losses reaches 0.5. For values of $\delta_\pm$ below about 0.65, $F^{-1}(0.01) = F^{-1}(0.5) < P_0 < F^{-1}(0.99)$ and $\Gamma(0.99) = +1$. The optimal prospect places more than half of the mass on losses and the skewness measure simplifies in a way which does not depend on either the location of this mass, nor on the location of the point $F^{-1}(0.99)$ describing the size of the right tail. Nonetheless, as probability weighting increases, the size of this right tail increases, even if this change cannot be captured in the skewness statistic.

To summarise, as the strength of probability weighting increases, the investor's return distribution changes from left or negatively skewed to right or positively skewed and the right tail becomes fatter. Most of this change is captured in the skewness statistic.

1.3.3 Recent experimental evidence

The vast majority of trading models - including those based on expected utility and those based purely on the S shaped utility function of prospect theory (Henderson (2012), Barberis and Xiong (2012), Ingersoll and Jin (2013) and Magnani (2016)) - predict investors sell stocks when the price breaches a threshold or pair of thresholds. For example, if an investor bought a stock at \$100, he should sell when the price rises to say \$105, or falls to say, \$90. If investors behaved according to such models, we should see them sell *the first time* the price breaches such threshold limits.

In fact, recent evidence of Viefers and Strack (2014) shows that this is very often not the case and individuals do not behave according to threshold rules. They conduct an experiment in a sophisticated asset selling task whereby subjects played sixty-five rounds during which they could sell their stock. In each round they observe a path of the market price which follows a random walk with positive drift. Viefers and Strack (2014) present evidence that players do not play cut-off or threshold strategies - they do not behave time-consistently within rounds 75% of the time, and visit the same price level three times on average before stopping at it. Their findings are supportive of our model whereby an investor with PT and probability weighting has an optimal target distribution over gains, because, as we will show in Section 1.5, our investor stops at a rate at each price level. Our research also has implications for the design of future experimental studies. Magnani (2016)'s recent experimental evidence to support the disposition effect is predicated on the behaviour of subjects being well approximated by threshold rules. Our theoretical findings, combined with Viefer and Strack's experimental observations, point to threshold-type behaviour providing an incomplete description of individuals' behaviours.

1.4 Explaining the disposition effect: the Odean's ratio

The disposition effect is arguably the most prominent trading anomaly in financial economics - the stylised fact that investors are on average more likely to sell a winner (an asset where the investor has a gain relative to purchase price) than a loser (where the investor has a loss). The effect has been documented for individual investors (Odean (1998)), institutional investors (Grinblatt and Keloharju (2001)) as well as in the real estate market (Genesove and Mayer (2001)) and options markets (Poteshman and Serbin (2003)). Studies have also examined the impact of trading experience (Feng and Seasholes (2005), Seru et al. (2010)) and investor sophistication (Dhar and Zhu (2006), Calvet et al. (2009), Grinblatt et al. (2012)) on the disposition effect. Finally, experimental evidence from the lab (Weber and Camerer (1998) and more recently, Magnani (2015) and Magnani (2016)) is also supportive.

In this and the next section we present two ways to measure the strength of the disposition effect within our theoretical model. The first is to construct a disposition measure based on the Odean (1998) measure, which focuses on the propensity to sell at a gain versus a loss without consideration of the size of those trades. More recently, researchers have studied how the sale propensity depends upon the magnitude as well as the sign of returns. Our second method presented in Section 1.5 is to develop a model-based implied sale intensity which can be compared to empirical selling schedules as recently documented in Barber and Odean (2013) and Ben-David and Hirshleifer (2012).

To test whether investors are disposed to selling winners and holding losers, we need to look at the frequency with which they sell winners and losers *relative* to their opportunities to sell each. Odean (1998) compares the proportion of gains realised (PGR) to the proportion of losses realised (PLR) by 10 000 individual investors with accounts at a discount brokerage firm over a six year period. Each time a stock is sold, the prices of all unsold stocks in the investors' portfolio are checked and it is recorded if they are trading at a gain, loss or neither on that day. The PGR (PLR) is the number of times a gain (loss) is realised as a fraction of the total number of times a gain (loss) could have been realised. Odean (1998) reports PGR=0.148 and PLR = 0.098, giving a disposition ratio of 1.51, or equivalently, investors realise gains at a 50% higher rate than losses. Using data over a different time period, Dhar and Zhu (2006) obtain a slightly higher ratio of 2.06.

Since we are working in continuous time, to capture the opportunities the investor had to sell at a gain (loss) we calculate the expected amount of time the price spent in the gain (loss) regime before a sale. A model-based measure of the rate of selling at gains (losses), denoted R_G (respectively, R_L) is found by dividing the probability of selling at a gain (loss) by the expected amount of time the price spent above (below) the initial price:

$$R_G = \frac{\mathbb{P}(P_\tau > P_0)}{\mathbb{E}(\int_0^\tau 1_{(P_u > P_0)} du)}, \qquad R_L = \frac{\mathbb{P}(P_\tau < P_0)}{\mathbb{E}(\int_0^\tau 1_{(P_u < P_0)} du)}$$

where τ is an optimal sale time in the model. Then, following Henderson (2012) (see also Magnani (2015)) we define the disposition ratio D by

$$D = \frac{R_G}{R_L} = \frac{\mathbb{P}(P_\tau > P_0)}{\mathbb{E}(\int_0^\tau 1_{(P_u > P_0)} du)} \frac{\mathbb{E}(\int_0^\tau 1_{(P_u < P_0)} du)}{\mathbb{P}(P_\tau < P_0)}. \tag{1.12}$$

This is the continuous time analog of Odean's measure. We say the disposition effect occurs when the ratio D is in excess of one.

Although the optimal prospect is typically unique there are several stopping rules or strategies which are optimal in the sense that they attain this optimal prospect. An important feature of (1.12) is that D does not depend on the stopping rule τ, except through the resulting prospect (the probability law of P_τ).

Proposition 1.8. *The disposition ratio D in* (1.12) *depends on the optimal prospect, but not on the stopping rule used to generate that prospect. Further, if the reference level R is equal to the initial price level P_0,*[16] *then D can be rewritten in the form*

$$D = \frac{\int_0^\infty \nu(dx)}{\int_0^\infty \frac{1}{\xi^2(x)} (u_\nu(x) - x)\, dx} \frac{\int_L^0 \frac{1}{\xi^2(x)} (u_\nu(x) + x)\, dx}{\int_L^0 \nu(dx)} \tag{1.13}$$

where ν is the probability law of the scaled target prospect $X_\tau = s(P_\tau)$, $u_\nu(x) := \mathbb{E}^{X\sim\nu}[|X - x|]$ is the potential function of ν and $\xi(x) := \sigma_P(s^{-1}(x))s'(s^{-1}(x))$.

Proof. By construction X satisfies the SDE $dX_t = \xi(X_t)dB_t$ with initial value $X_0 = 0$. Let $L^X = (L_t^X(x))_{t\geq 0, s(a_J)\leq x\leq s(b_J)}$ be the local time of X at level x by time t using the standard normalisation of, say, Revuz and Yor (1999). (No confusion should arise between $L = s(a_J)$ and the local time $L_t^X(x)$ since the former never has any sub- or superscripts.) By the occupation times formula (Revuz and Yor (1999) [Theorem VI.1.6]), for any Borel function Φ,

$$\int \Phi(a)L_t^X(a)da = \int_0^t \Phi(X_s)d[X]_s. \tag{1.14}$$

[16]The choice of the reference level is not crucial and we impose $R = P_0$ purely for the purpose of expressions simplification.

Clearly, $\mathbb{P}(P_\tau > P_0) = \mathbb{P}(X_\tau > 0) = \int_0^\infty \nu(dx)$. Similarly $\mathbb{P}(P_\tau < P_0) = \int_L^0 \nu(dx)$. On the other hand,

$$\begin{aligned}
\mathbb{E}\left(\int_0^\tau 1_{(P_u>P_0)}du\right) = \mathbb{E}\left(\int_0^\tau 1_{(X_u>0)}du\right) &= \mathbb{E}\left(\int_0^\tau \frac{1_{(X_u>0)}}{\xi^2(X_u)}d[X]_u\right) \\
&= \mathbb{E}\left(\int \frac{1_{(x>0)}}{\xi^2(x)}L_\tau^X(x)dx\right) \\
&= \int_0^\infty \frac{\mathbb{E}(L_\tau^X(x))}{\xi^2(x)}dx
\end{aligned}$$

where we use (1.14) for the penultimate equality. But, by Tanaka's formula $\mathbb{E}(L_\tau^X(a)) = \mathbb{E}|X_\tau - a| - |X_0 - a|$. Hence, writing $u_\nu(x) := \int |z-x|\nu(dz)$

$$\mathbb{E}\left(\int_0^\tau 1_{(P_u>P_0)}du\right) = \int_0^\infty \frac{1}{\xi^2(x)}\left(\mathbb{E}|X_\tau - x| - |X_0 - x|\right)dx = \int_0^\infty \frac{1}{\xi^2(x)}\left(u_\nu(x) - x\right)dx$$

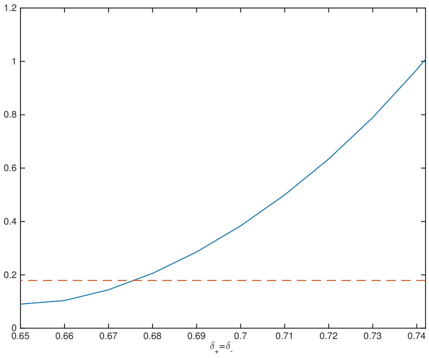
which is independent of the stopping rule used to realise ν. Similarly we can establish

$$\mathbb{E}\left(\int_0^\tau 1_{(P_u<P_0)}du\right) = \int_L^0 \frac{1}{\xi^2(x)}\left(u_\nu(x) + x\right)dx.$$

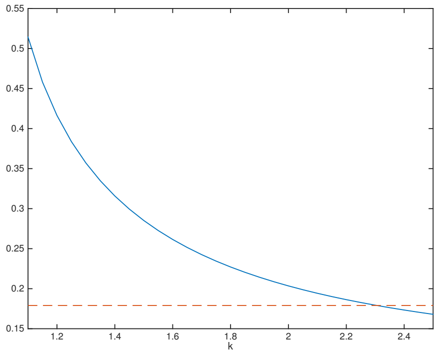
The representation (1.13) follows. □

If we assume the investor is a PT value maximiser, then his optimal scaled prospect can be computed from the optimal quantile function (see (1.11)), and we can calculate D without making any assumption about how the investor trades to achieve the optimal distribution.

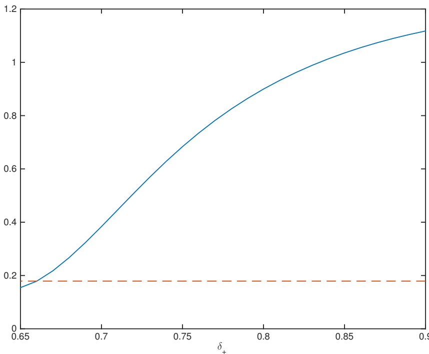
Figure 1.5 plots the (base-10 logarithm of the) disposition ratio D against the Tversky and Kahneman (1992) weighting parameter in panel (a), loss aversion in panel (b), and the weighting parameters separately in panels (c) and (d). Other parameters are our base values. The horizontal dashed line in each panel represents Odean's PGR/PLR of $\log_{10} 1.51 \approx 0.18$. Our main finding is that in contrast to PT models without probability weighting, the model incorporating weighting can indeed deliver Odean's estimate. In panel (a), we see that for a probability weighting parameter $\delta_\pm$ of about 0.675, we obtain a disposition ratio of about 1.5. In panel (b), holding other parameters fixed and varying loss aversion, we see that for a loss aversion of around 2.25, the disposition ratio is again about 1.5. The lower two panels (c) and (d) look at varying the two probability weighting parameters separately. In panel (d) we see that values close to Odean's measure can be generated by a range of δ_- between about 0.5 and 0.65, when $\delta_+ = 0.7$. It is also worth noting that Dhar and Zhu (2006)'s disposition measure of 2.06 ($\log_{10} 2.06 \approx 0.313$) can be obtained with slightly less probability weighting and a lower level of loss aversion.



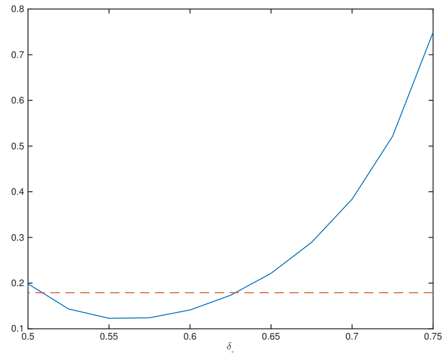
(a) $\delta_+ = \delta_-$ varying



(b) k varying



(c) δ_+ varying



(d) δ_- varying

Figure 1.5: Base-10 logarithm of the disposition ratio D given in (1.12). Each panel varies one parameter, keeping others fixed at base values. Other base parameters used are $\delta_\pm = 0.7$, $\alpha_+ = 0.5$, $\alpha_- = 0.9$, $\beta = 0.9$, $k = 1.25$. The reference level is $R = 1$, and the current price is $P_0 = 1$. The horizontal dashed lines mark Odean's disposition estimate of $\log_{10} 1.51 \approx 0.18$.

We see that as the degree of probability weighting becomes small, the measure of the disposition effect gets very large. As $\delta_\pm \to 1$, we recover the case in the absence of probability weighting studied in Henderson (2012) where the calibrated disposition measure was much greater than that found in the empirical data. Thus, probability weighting does indeed help PT explain realistic levels of the disposition effect.

We also comment on the impact of the level of risk aversion and risk seeking on the disposition ratio. Parameters around the base values deliver Odean's ratio of about 1.5. If the investor is more risk averse over gains, then he will shrink the range of price gains over which he stops, so will take smaller gains more frequently whilst also taking fewer losses, which increases the disposition ratio. In the extreme, as $1 - \alpha_+$ gets sufficiently large, the investor waits indefinitely over losses (see Figure 1.3) and the ratio is infinite. If the investor is more risk seeking over losses, then the increased convexity in the loss region drives the disposition ratio higher as the investor delays taking losses. In the extreme as $1 - \alpha_-$ is sufficiently large, the investor waits indefinitely over losses (see the thresholds in Figure 1.3 (d)) and the disposition ratio becomes infinite.

It is worth highlighting that the model has delivered Odean's estimate of the disposition effect even for a single investor - at this point we have not needed to extend to a mixture over heterogeneous investors.[17] In contrast, in the setting of Ingersoll and Jin (2013) without probability weighting but allowing for reinvestment, heterogeneity was necessary to obtain a fit with Odean's measure. They mix reference-dependent realisation utility traders with random Poisson traders in a 50-50 ratio to obtain a good fit to the empirical data. Here, in our model, we instead have the impact of probability weighting, which is working to enable PT to deliver realistic levels of the disposition effect.

1.5 Explaining the disposition effect: implied selling intensity

In this section, we extend our analysis to consider the relative *magnitude* of gains and losses. The disposition effect is commonly understood as the preference for selling assets that have increased in value relative to assets that have decreased in value since purchase. Researchers have recently studied how the rate of sale depends on the relative magnitude of the gain or loss. A number of authors estimate proportional hazards models to derive the hazard rate

[17] Although such an extension is straightforward and would clearly also be able to achieve realistic levels of the disposition ratio, it is not necessary for a good fit.

for the sale of stock conditional on return since purchase, see Feng and Seasholes (2005), Seru et al. (2010) and Barber and Odean (2013). Others, in particular, Ben-David and Hirshleifer (2012) (see also Kaustia (2010)) document the probability of selling as a function of profits, whilst allowing for different prior holding periods to be taken into account.

There is broad agreement amongst researchers that the estimated hazard rate as a function of returns since purchase is higher on gains than on losses. This is an evidence in favour of a disposition effect amongst investors because the higher rate over gains means that the average propensity to sell is higher for gains than losses. For all but very short holding periods, researchers consistently find the hazard rate or selling schedule on losses is fairly flat (see Ben-David and Hirshleifer (2012), Barber and Odean (2013) and Seru et al. (2010)).

There is less consensus over results concerning the overall shape of the hazard rate or selling schedule in the literature with findings depending upon the length of the holding period under consideration. For instance, over short holding periods, Ben-David and Hirshleifer (2012) demonstrate a strong asymmetric V shaped pattern in their empirical selling schedules. Some authors even find that when holding periods are aggregated, the selling intensity function may exhibit an inverted V shape (Odean (1998), Meng and Weng (2016)).

In the remainder of this section, we will develop a model-based selling intensity for a single and for many investors, and compare to the findings of the empirical literature. Since our model gives an intensity over all holding periods, our focus is on achieving the empirical features coming from the aggregated data.

1.5.1 Model-based implied selling intensity

The empirical measure of the selling rate at price level p can be defined as

$$\gamma(p) = \frac{\text{number of sales at } p}{\text{amount of time spent at } p}.$$

The equivalent model-based quantity is

$$\zeta(p) = \frac{\mathbb{P}(P_\tau \in dp)}{\mathbb{E}\left[\int_0^\tau du 1_{(P_u \in [p,p+dp))}\right]} \tag{1.15}$$

provided this quantity is well defined. Similar to our model-based disposition ratio in (1.12), it can be shown that $\zeta(p)$ is the same for *all* stopping times τ such that P_τ has law μ.

Proposition 1.9. *Suppose μ is the law of the target prospect P_τ and ν is the law of the scaled prospect $X_\tau = s(P_\tau)$. Assume a reference point of $R = P_0$. If μ has a density χ then*

$$\zeta(p) = \frac{\sigma_P(p)^2 s'(p)}{u_\nu(s(p)) - |s(p)|}\chi(p). \tag{1.16}$$

Proof. Recall the notation we use in the proof of Proposition 1.8. For any arbitrary Borel function Φ we have

$$\int_0^\tau \Phi(P_u)d[P]_u = \int_0^\tau \Phi(s^{-1}(X_u))\frac{\sigma_P^2(s^{-1}(X_u))}{\xi^2(X_u)}d[X]_u = \int L_\tau^X(a)\Phi(s^{-1}(a))\frac{da}{[s'(s^{-1}(a))]^2}$$
$$= \int L_\tau^X(s(u))\frac{\Phi(u)}{s'(u)}du.$$

In particular if we choose $\Phi(z) = \frac{1_{(z\in[p,p+dp))}}{\sigma_P^2(z)}$ then we have

$$\int_0^\tau 1_{(P_u\in[p,p+dp))}du = \frac{L_\tau^X(s(p))}{\sigma_P^2(p)s'(p)}dp.$$

The expected value of the above expression can be computed by Tanaka's formula, and (1.16) follows.

□

More generally, to allow for optimal prospects which contain atoms we set

$$\lambda(dp) = \frac{\sigma_P(p)^2 s'(p)}{u_\nu(s(p)) - |s(p)|}\mu(dp). \tag{1.17}$$

If μ is absolutely continuous then $\lambda(dp) = \zeta(p)dp$. Conversely, if the optimal selling rule is a pure threshold strategy, i.e. a strategy in which it is optimal to sell the asset the first time the price process leaves an interval (whence the optimal prospect is a pair of point masses) then we find $u_\nu(s(p)) = |s(p)|$ at the ends of the interval and the measure λ consists of a pair of point masses of infinite size. (This is intuitive: we must stop the price process at the first time it leaves the relevant interval, and the only way we can ensure this is to stop at an infinite rate.) We have seen an example of this when there is no probability weighting in the model.

In fact, the optimal selling rule in our asset liquidation model contains a point mass on losses, and a mixture distribution on gains consisting of a point mass and a continuous distribution above that point. The corresponding stopping rate λ has an atom of infinite size at the location of the point mass on losses, an atom of finite size at the location of the point mass on gains, and a continuous density above this point.

Given a non-negative function $g = g(p)$ then we can consider selling at a rate $g(P_t)$ per unit time. This is equivalent to stopping at the random time $\tau^g = \inf\{u : \int_0^u g(P_v)dv > T\}$ where T is an independent exponential random variable with unit parameter. Given a target law, for example the law μ^* of the optimal prospect, we can ask if it is possible to choose a level dependent (but not explicitly time-dependent) function g such that P_{τ^g} has the desired target law. This can be done, and makes use of the measure λ in (1.17), as described in the following lemma where the proof is given in Appendix 1.E.

Lemma 1.10. *Let $\Lambda^P = (\Lambda^P_t)_{t\geq 0}$ be the increasing additive functional $\Lambda^P_t = \int \ell^P_t(p)\lambda(dp)$, where $\ell^P = (\ell^P_t(p))_{p>0,t\geq 0}$ is an appropriately scaled version of local time process of P.*

If T is an independent, exponentially distributed, unit rate random variable and if $\tau = \inf\{u : \Lambda^P_u > T\}$, then $P_\tau \sim \mu^*$.

This lemma gives a second interpretation of the quantity $\lambda(dp)$: if the investor sells at a level-dependent rate per unit time given by λ, then he will attain the optimal prospect. Note that we are not arguing that our investors must follow this stopping rule, but rather this kind of randomised strategy provides a convenient way to interpret the model-based quantities given in (1.15) and (1.17).

1.5.2 Mixing over heterogeneous investors

We have so far discussed the case of a single investor implementing a stopping rule to generate a target prospect. However, the typical empirical selling rate estimated from market data is an amalgamation of liquidation strategies enacted contemporaneously by multiple investors who may have different risk preferences. We consider the implied selling intensity function when we average across individuals.

Let Θ denote the space of risk and probability weighting parameters. If the parameters of the typical investor are distributed with prior law η on Θ, then we find a model-based selling rate at price level p of

$$\overline{\zeta}(p)dp = \overline{\lambda}(dp) = \frac{\overline{\mu}(dp)\sigma_P(p)^2 s'(p)}{u_{\overline{\nu}}(s(p)) - |s(p)|} \tag{1.18}$$

where $\overline{\mu}(dp) = \int \eta(d\theta)\mu_\theta(dp)$ and $\overline{\nu}$ is given by $F_{\overline{\nu}}(x) = F_{\overline{\mu}}(s^{-1}(x))$.

To illustrate the idea, we suppose the price process P follows a geometric Brownian motion and the expected return is such that P is a martingale. We are free to choose Θ the space of parameters and η, the distribution over this parameter space. For this example, we suppose investors have a common pair of probability weighting functions (given by TK with $\delta_\pm = 0.7$) and TK value functions (with reference level $R = P_0 = 1$). We assume $\alpha_+ = 0.5$ and $\alpha_- = 0.9$ are fixed across investors but that the loss aversion parameter k varies. We identify Θ with a subset of $\mathbb{R}_+$ corresponding to the value of the loss aversion parameter. Once we have specified η we can calculate the implied selling density. Different choices of η will lead to different model-based predictions for the selling density. We make use of the earlier empirical observation that the selling rate on losses is approximately constant across different returns. We design η such that the model-based selling intensity at a loss is constant, and then consider the implications for the model-based rate of selling at gains. The construction of such a prior density is given in Appendix 1.F. Barber and Odean (2013) find hazard ratios of around one for losses in their analysis of the LDB and Finnish datasets. This motivates our choice of a unit rate of selling on losses, which corresponds to a daily probability of a loss of 1/250 or 0.4%.

Figure 1.6 displays our model's implied stopping rate $\overline{\zeta}(p)$ against price. The model implied sales rate on gains, given in (1.18) is plotted. Our first goal is to demonstrate that the disposition effect holds. The implied sales rates are indeed consistent with the disposition effect as the rate for gains is higher than that for losses and thus implies a higher propensity to sell at gains than losses. The second goal is to show the model implied sales rate captures some of the features of the empirical data. Since our focus is on results for longer holding periods or aggregate data, we compare with the estimated hazard ratios of Barber and Odean (2013) and the graphs of the probability of selling shares at Day 60, 125, and 250 in Figure 1 of Ben-David and Hirshleifer (2012).

In particular, the graphs in Figure 1 of Ben-David and Hirshleifer (2012) for Day 60 and 250 have very similar qualitative features to our Figure 1.6. Over losses, the probability of sale is relatively flat. Over gains, the sales probability is slightly humped - first rising and then falling with the magnitude of returns since purchase. This is also true (perhaps to a lesser extent) in the graph of the hazard ratio for the Finnish dataset in Barber and Odean (2013).

The rate of $\overline{\zeta}$ between 3.5-5 in Figure 1.6 equates to a daily probability of selling between 1.4% and 2%. The implied sales rate in our example is slightly higher than those in Barber and Odean (2013) and Ben-David and Hirshleifer (2012) and thus our disposition effect is on the strong side.

The presence of probability weighting in our model has had a dramatic impact on the PT model's ability to produce realistic sale intensity schedules across different return magnitudes. Without probability weighting, recall that the optimal distribution for an individual is a two point distribution with weight on a single gain and a single loss threshold. Mixing over investors can improve the fit, but, as Ingersoll and Jin (2013) find, in the absence of probability weighting, PT investors need to be mixed with random Poisson traders to yield a reasonable calibration.

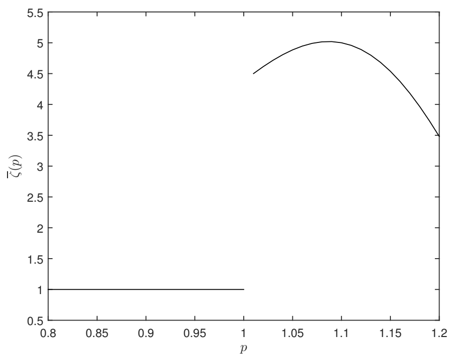


Figure 1.6: Implied stopping rate $\overline{\zeta}(p)$ against price for a set of heterogeneous investors. Parameters are $\delta_{\pm} = 0.7$, $\alpha_+ = 0.5$, $\alpha_- = 0.9$, $\sigma = 0.5$. Loss aversion k varies across investors. The reference level is $R = P_0 = 1$.

1.6 Conclusion

Despite the recent surge in interest in the implications of prospect theory in dynamic trading models, very few papers treat the probability weighting element of prospect theory, in part because of the additional technical challenges probability weighting brings. In this chapter we argue the understanding of the implications of probability weighting in a dynamic setting is crucial, as it significantly alters the form of the PT investor's optimal strategy.

The behaviour of PT investors in the absence of probability weighting has been studied by Ingersoll and Jin (2013), Henderson (2012) and Barberis and Xiong (2012). They find that the optimal stopping rule in this setting is a stop-loss, stop-gain strategy. In fact, although there is some support for stop-loss strategies in empirical and laboratory data, stop-gain strategies are not widely supported. Moreover, the locations of the stop-loss and stop-gain thresholds are such that a gain is much more likely than a loss, corresponding to a target law with a negative skew and an extreme disposition effect (well beyond that found in empirical data). Further, it is not possible to obtain reasonable patterns for selling rates at different price levels even when mixing over PT agents with characteristics based on different parameter values.

We show that re-introducing probability weighting improves the predictive power of models of PT agents in all these aspects. With probability weighting set to levels estimated in the literature we find the PT investor has an optimal stopping rule which is stop-loss but not stop-gain. Instead, the investor trades to achieve an optimal prospect which on gains is a long-tailed distribution chosen to reflect the overweighting of extreme gains. Overall, his target prospect can have right skew, and a model-based disposition ratio which matches the levels predicted by Odean (1998). Finally, by considering mixtures of PT agents we find model-based level-dependent selling intensities which closely mirror the patterns of level-dependent selling rates which have been documented in the literature. Both the model-based rate and the empirical rate are approximately constant on the loss regime, and much higher on gains, consistent with the disposition effect.

Appendix to Chapter 1

1.A Existing results on optimal stopping with probability weighting

Our goal in this and the next section is to prove Proposition 1.7 which characterises the optimal prospect. The starting point of our results is Xu and Zhou (2013) who have the important insight that the optimisation over stopping times can be reduced to a problem of maximising over distributions. However, their results are largely confined to the one-sided case of gains (or losses) only. We summarise the relevant results in this section.

Recall that the quantile function of a random variable Y, denoted G_Y (or G if the random variable is clear), is the (left-continuous) inverse of the cumulative distribution function.

Proposition 1.11 (Lemmas 3.1 and 3.2 of Xu and Zhou (2013))**.** *Suppose the scaled asset price process $X = (X_t)_{t\geqslant 0}$ is always non-negative and $X_0 > 0$. Then the probability-weighted optimal stopping problem*

$$\sup_{\tau} \int_0^{\infty} w\left(\mathbb{P}\left(v(X_\tau) > x\right)\right) dx \tag{1.19}$$

has a dual representation in terms of the quantile function $G = G_X$ of X_τ in the form

$$\sup_{G \in A_{X_0}} \int_0^1 v(G(x)) w'(1-x) dx \tag{1.20}$$

where

$$A_z = \{G|\ G : (0,1) \to [0,\infty) \text{ is a left-continuous quantile function, } \int_0^1 G(x)dx = z\}.$$

In particular, if G^* is an optimiser of (1.20) *with ν^* being the associated probability law, then there exists a stopping time τ^* such that $X_{\tau^*} \sim \nu^*$, and such τ^* is optimal for* (1.19)*.*

The optimal prospects in this one-sided problem can be identified under some particular forms of v and w. The two results below are the most relevant to our current problem.

Proposition 1.12 (Theorem 5.2 and Lemma 4.1 of Xu and Zhou (2013))**.** *The optimiser of* (1.20) *can be characterised under the two cases below:*

1. *Suppose the target prospect can take values on* $[0, \infty)$*, and that the mean is constrained to be less than or equal to* z*. If* v *is concave and* w *is inverse-S shaped such that it is concave on* $[0, q]$ *and convex on* $[q, 1]$*, then the optimiser* G *is of the form*
$$G(x) = a1_{(0,1-q]} + \left(a \vee (v')^{-1}\left(\frac{\lambda}{w'(1-x)}\right)\right)1_{(1-q,1)} \tag{1.21}$$
for some $a \geqslant 0$ *and* $\lambda \geqslant 0$*, where* a *and* λ *are chosen such that they respect the constraint* $\int_0^1 G(x)dx = z$*. The optimal prospect is an atom at* a *of size at least* $1 - q$ *combined with a continuous distribution on* $(a \vee (v')^{-1}(\frac{\lambda}{w'(q)}), \infty)$*.*

2. *Suppose the target prospect is bounded such that it can only take values on* $[0, K]$ *for some* $0 < K < \infty$*, and its mean is constrained to be less than or equal to* z*. If* v *is convex and* w *is a general probability weighting function, the optimiser* G *is a step-function taking values on* 0*,* K *and some* $b \in (0, K)$*. The optimal prospect is a three-point distribution with masses at* 0*,* b *and* K*.*

Proof. In the first case with a concave v, one can write down the relaxed Lagrangian objective function and obtain a candidate optimiser as $(v')^{-1}\left(\frac{\lambda}{w'(1-x)}\right)$ for some Lagrangian multiplier λ from the first order condition. This candidate optimiser is not monotonic since w is inverse-S shaped and thus it cannot be a quantile function. But it can be shown that the optimal solution is in form of a suitably truncated version of this candidate optimiser. See Xu and Zhou (2013) for details.

The second case involving a convex v is much more subtle since it is a constrained convex maximisation problem on a somewhat complicated space. Xu and Zhou (2013) provide a proof[18] based on an approximating sequence with step functions. Here we give an alternative proof using a more fundamental approach.

Suppose G is the optimiser to the problem in the second case and the image of G contains more than three distinct elements. Then we can choose $(c, b) \in (0, 1) \times (0, K)$ with $b \in [G(c), G(c+)]$ such there exists $x_1 \in (0, c)$ and $x_2 \in (c, 1)$ with $0 < G(x_1) < b$ and $b < G(x_2) < K$.

Consider a convex function $\overline{\eta}_1(x) = \overline{\eta}_1(x; \epsilon_1) := x^{\epsilon_1}$ and a concave function $\overline{\eta}_2(x) = \overline{\eta}_2(x; \epsilon_2) := 1 - (1 - x)^{\epsilon_2}$ with $\epsilon_1 > 1$ and $\epsilon_2 > 1$. Define a new quantile function via
$$\overline{G}(x) = \overline{G}(x; \epsilon_1, \epsilon_2) := \begin{cases} b\overline{\eta}_1\left(\frac{G(x)}{b}\right), & 0 < x \leqslant c, \\ b + (K - b)\overline{\eta}_2\left(\frac{G(x)-b}{K-b}\right), & c < x < 1. \end{cases}$$

[18] Although Xu and Zhou (2013) do not directly consider the problem with bounded payoff, their Lemma 4.1 can be trivially extended to a set of quantile functions with bounded range on $[0, K]$.

With this definition we have $\overline{G}(x) \leqslant G(x)$ on $(0, c]$ and $\overline{G}(x) \geqslant G(x)$ on $(c, 1)$. Moreover, there exists values of x such that these inequalities are strict.

Then, $\triangle_1(\epsilon_1) := \int_0^c G(x)dx - \int_0^c \overline{G}(x)dx > 0$ and $\triangle_2(\epsilon_2) := \int_c^1 \overline{G}(x)dx - \int_c^1 G(x)dx > 0$. Observe that $\triangle_1$ and $\triangle_2$ are strictly increasing in ϵ_1 and ϵ_2 respectively, and $\triangle_1, \triangle_2 \downarrow 0$ when $\epsilon_1, \epsilon_2 \downarrow 1$. Thus we can choose $\epsilon_1^* > 1$ and $\epsilon_2^* > 1$ such that $\triangle_1(\epsilon_1^*) = \triangle_2(\epsilon_2^*)$, or equivalently $\int_0^1 \overline{G}(x; \epsilon_1^*, \epsilon_2^*)dx = \int_0^1 G(x)dx \leqslant z$.

Now fix $0 < \lambda^* < \min\left(\frac{1}{\epsilon_1^*}, \frac{1}{\epsilon_2^*}\right) < 1$. Consider another pair of functions given by

$$\widetilde{\eta}_i(x) := \frac{x - \lambda^* \overline{\eta}_i(x; \epsilon_i^*)}{1 - \lambda^*}$$

for $i = 1, 2$. It can be easily checked that $\widetilde{\eta}_i(0) = 0$ and $\widetilde{\eta}_i(1) = 1$, and $\widetilde{\eta}_1$ (resp. $\widetilde{\eta}_2$) is a strictly increasing concave (resp. convex) function on $[0, 1]$.

Define another quantile function via

$$\widetilde{G}(x; \epsilon_1^*, \epsilon_2^*) := \begin{cases} b\widetilde{\eta}_1\left(\frac{G(x)}{b}\right), & 0 < x \leqslant c, \\ b + (K - b)\widetilde{\eta}_2\left(\frac{G(x) - b}{K - b}\right), & c < x < 1. \end{cases}$$

Then from the construction of $\widetilde{\eta}$ it can be easily verified that $\int_0^1 \widetilde{G}(x; \epsilon_1^*, \epsilon_2^*)dx = \int_0^1 \overline{G}(x; \epsilon_1^*, \epsilon_2^*)dx \leqslant z$ and

$$G(x) = \lambda^* \overline{G}(x; \epsilon_1^*, \epsilon_2^*) + (1 - \lambda^*)\widetilde{G}(x; \epsilon_1^*, \epsilon_2^*). \tag{1.22}$$

Convexity of v now leads to

$$\int_0^1 v(G(x))w'(1 - x)dx \leqslant \max\left(\int_0^1 v(\overline{G}(x))w'(1 - x)dx, \int_0^1 v(\widetilde{G}(x))w'(1 - x)dx\right). \tag{1.23}$$

But both $\overline{G}$ and $\widetilde{G}$ are feasible solutions to the optimisation problem. The strict convexity of v and the optimality of G implies equality has to hold in (1.23) and in turn in (1.22) which happens if and only if $G(x) = \overline{G}(x) = \widetilde{G}(x)$. This means

$$G(x) = \begin{cases} b\overline{\eta}_1\left(\frac{G(x)}{b}\right), & 0 < x \leqslant c, \\ b + (K - b)\overline{\eta}_2\left(\frac{G(x) - b}{K - b}\right), & c < x < 1. \end{cases}$$

or equivalently

$$\begin{cases} \frac{G(x)}{b} = \overline{\eta}_1\left(\frac{G(x)}{b}\right), & 0 < x \leqslant c, \\ \frac{G(x) - b}{K - b} = \overline{\eta}_2\left(\frac{G(x) - b}{K - b}\right), & c < x < 1. \end{cases} \tag{1.24}$$

Given $\epsilon_1^* > 1$ and $\epsilon_2^* > 1$, (1.24) implies $\frac{G(x)}{b}$ must take values on 0 or 1 over $(0, c]$, and $\frac{G(x) - b}{K - b}$ must take values on 0 or 1 over $(c, 1)$. Hence $G(x)$ can only take values on 0 or

b over $(0, c]$, and b or K over $(c, 1)$. This contradicts the assumption that the image of G contains more than three distinct elements. Hence the optimal quantile function must be a three-step function, and moreover it can only take values on 0, K and some interim level $b \in (0, K)$.

$\square$

The results in the second part of the Proposition 1.12 translate directly to a one-sided problem involving losses. Thus, we can deduce from the results of Xu and Zhou (2013) that in the loss regime the optimal prospect consists of up to three point masses.

Our main technical contribution is to solve for the optimal prospect for a wide class of prospect theory specifications. The results we present are not valid for all set-ups. Rather, we make some additional assumptions involving our elasticity condition which are satisfied under our base case, and more widely under many standard formulations of the problem. Under these additional assumptions we can prove that the optimal prospect has extra structure beyond that which can be deduced from Proposition 1.12. This extra structure allows us to solve for the optimal prospect in the general case with probability weighting and both gains and losses.

First, on the gain regime we show that a and λ are such that $a \geq (v')^{-1}(\frac{\lambda}{w'(q)})$. Hence the point a is simultaneously the location of a point mass in the optimal prospect and the lower limit in the continuous part of the optimal prospect on gains. Second, on the loss regime, we show that the optimal prospect is a single point mass (and not three point masses) located at some point $b \in [L, 0)$. These results are shown in the next section.

1.B General construction of the optimal solution

In this section, we solve (1.7) assuming that the scaled value function v is concave on the gain regime $[0, \infty)$ and convex on the loss regime $[L, 0]$. The probability weighting functions $w_\pm$ are inverse-S shaped, concave on $[0, q_\pm]$ and convex on $[q_\pm, 1]$. The starting level of the scaled price process X_0 is a given fixed constant. Our base case fits into this setting.

Using the same ideas as in Proposition 1.11, (1.7) is equivalent to

$$\sup_X \left(\int_{1-\mathbb{P}(X>0)}^{1} v(G_X(u)) w'_+(1-u) du + \int_0^{\mathbb{P}(X<0)} v(G_X(u)) w'_-(u) du \right), \tag{1.25}$$

where the supremum is taken over random variables (or prospects) X with mean X_0 and support on $[L, \infty)$. If X^* is an optimal prospect, then any stopping time τ^* constructed such that X_{τ^*} has the same law as X^* is optimal.

1.B.1 The problem for gains

Suppose we are given $\phi_+ \in (0,1]$ and $z_+ \geqslant X_0^+$ which are the probability mass allocated to gains and the mean of gains. The gain problem is to find

$$D_+(\phi_+, z_+) = \sup_{G \in A^+_{\phi_+, z_+}} \int_{1-\phi_+}^1 v(G(x)) w'_+(1-x) dx \tag{1.26}$$

where

$$A^+_{\phi,z} = \{G |\ G : (0,1) \to [0,\infty) \text{ is a quantile function, } \int_{1-\phi}^1 G(x)dx = z,\ G(x) = 0 \text{ on } (0, 1-\phi]\}.$$

The gain problem involves an optimisation for concave v and inverse-S shaped w_+. The first part of Proposition 1.12 can be applied to identify the form of the optimal quantile function on $(1-\phi_+, 1)$.[19] From (1.21), with q_+ being the point of inflexion of w_+ we deduce that the optimiser is of the form

$$\begin{aligned} G_+(x) = G_+(x; \phi_+, z_+; a, \lambda) = a 1_{(1-\phi_+, (1-q_+)\vee(1-\phi_+)]} \\ + \left(a \vee (v')^{-1} \left(\frac{\lambda}{w'_+(1-x)} \right) \right) 1_{((1-q_+)\vee(1-\phi_+), 1)} \end{aligned} \tag{1.27}$$

for some constants $a \geqslant 0$ and $\lambda \geqslant 0$. (Note that $G_+ = 0$ on $(0, 1-\phi_+]$.) The optimal values a^* and λ^* of a and λ are obtained by maximising the objective function in (1.26) over $A^+_{\phi_+, z_+}$.

Lemma 1.13. *Suppose v' is continuous with $v'(0+) = \infty$ and $\lim_{x\uparrow\infty} v'(x) = 0$. Then for the optimal prospect we have $a^* \geq (v')^{-1} \left(\frac{\lambda^*}{w'_+(q_+ \wedge \phi_+)} \right)$.*

Proof. We begin by showing that the optimiser (1.21) to the one-sided gain-problem (1.20) with v concave has its parameter $a \geq (v')^{-1}(\frac{\lambda}{w'_+(q_+)})$.

Let $L(x; G, \lambda) = v(G) w'_+(1-x) - \lambda G$. This is maximised over G by $G = y_L(x; \lambda) := (v')^{-1} \left(\frac{\lambda}{w'_+(1-x)} \right)$. If $v'(0) = \infty$ and $\lim_{x\uparrow\infty} v'(x) = 0$, then $y_L(x; \lambda)$ is well defined and strictly positive for all $0 < x < 1$. Now suppose G_0 is an optimal solution to (1.20) which has the form of (1.21) with parameters (a_0, λ_0). Consider another quantile function G_1 which is in form of (1.21) with parameters (a_1, λ_1) where $a_1 := (v')^{-1} \left(\frac{\lambda_1}{w'_+(q_+)} \right) > a_0$ and the value of λ_1 is implied by the constraint $\int_0^1 G_1(x)dx = z$.

[19] By considering a transformation of $\widetilde{G}(x) = G((1-\phi_+)(1-x) + x)$, one can see (1.26) can be rewritten in form of (1.20).

On $0 < x \leqslant 1 - q_+$, $G_0(x) = a_0 < a_1 = G_1(x) \leqslant y_L(x;\lambda_1)$. As $y_L(\,\cdot\,;\lambda_1)$ is the maximiser of $L(\,\cdot\,;G,\lambda_1)$ we have $L(x;G_1,\lambda_1) \geqslant L(x;G_0,\lambda_1)$. On $1 - q_+ < x < 1$, $G_1(x) = y_L(x;\lambda_1)$ and then trivially $L(x;G_1,\lambda_1) \geqslant L(x;G_0,\lambda_1)$. This shows $L(x;G_1,\lambda_1) \geqslant L(x;G_0,\lambda_1)$ for all $0 < x < 1$ (with strict inequality holding for some x). It contradicts the assumption that G_0 is an optimal solution.

By extension, the optimiser (1.27) for the sub-problem of gains (1.26) must have its parameter $a \geq (v')^{-1}\left(\frac{\lambda}{w'_+(q_+\wedge\phi_+)}\right)$. □

Two things follow. First, since $a > 0$ an optimal solution on gains must allocate all the available probability mass on gains. This also implies the value of the gain problem $D_+(\phi_+, z)$ is always strictly increasing in the available probability mass ϕ_+.

Second, $a = (v')^{-1}\left(\frac{\lambda}{w'_+(\psi)}\right)$ for some $\psi \leq q_+ \wedge \phi_+$ and then we can rewrite the optimal solution in (1.27) as

$$G_+(x) = G_+(x;\phi_+,\psi,\lambda) = (v')^{-1}\left(\frac{\lambda}{w'_+(\psi\wedge(1-x))}\right)1_{(1-\phi_+,1)} \tag{1.28}$$

where the mean on gains can be written as

$$z_+ = z_+(\phi_+,\psi,\lambda) = \int_{1-\phi_+}^{1} (v')^{-1}\left(\frac{\lambda}{w'_+((1-u)\wedge\psi)}\right)du = \int_0^{\phi_+} (v')^{-1}\left(\frac{\lambda}{w'_+(u\wedge\psi)}\right)du. \tag{1.29}$$

Then G_+ is identically 0 on $(0, 1-\phi_+)$, equal to a constant on $(1-\phi_+, 1-\psi)$ and continuous on $(1-\phi_+, 1)$. The corresponding distribution has an atom at some a^* and a density on (a^*, ∞). The candidate optimiser takes the form in (1.28) parameterised by ϕ_+, λ and $\psi \leq q_+ \wedge \phi_+$ subjected to (1.29).

1.B.2 The problem for losses

Suppose now we are given ϕ_- and z_- such that $\phi_- \geqslant 0$ and $X_0^- \leqslant z_- \leqslant K\phi_-$ where $K = -L$ and where ϕ_- and z_- represent the probability mass allocated to losses and the mean of losses. The loss problem is to find

$$D_-(\phi_-, z_-) = \sup_{G\in A^-_{\phi_-,z_-}} \int_0^{\phi_-} v(G(x))w'_-(x)dx \tag{1.30}$$

where

$$A^-_{\phi,z} = \{G|\ G:(0,1)\to[L,0] \text{ is a quantile function}, \int_0^{\phi} G(x)dx = -z,\ G(x) = 0 \text{ on } (\phi,1)\}.$$

The loss problem is a maximisation problem involving a convex v and an inverse-S shaped w_- over quantile functions with fixed mean and a bounded range. Using the second part of Proposition 1.12, the optimal quantile function on $(0, \phi_-)$ is in the form of a step-function taking values on L, 0 and some $b \in (L, 0)$. Our main result is that under some mild extra assumptions on v and w_- the solution can be further simplified to a two-point distribution which has a mass at L or a mass at an interior point, but not both, together with an atom at 0. Later we will argue that for the problem with gains and losses there cannot be a mass at zero.

Proposition 1.14. *If $E(x; v, L)$ is increasing in $x \in [L, 0]$ and $E(p; w_-, c)$ is decreasing in $p \in [0, \min(q_-, c)]$ for any $c \in [0, 1]$, then the optimal solution to problem (1.30) is a two-point distribution with probability mass allocated to 0 and a single further point in $[L, 0)$. The optimal quantile function is of the form*

$$G_-(x; \phi_-, z_-) = -\frac{z_-}{\eta} 1_{(0,\eta]}$$

for some η with $-\frac{z_-}{L} \leqslant \eta \leqslant \phi_-$.

Proof. Using the second part of Proposition 1.12 we know that on losses, the optimal prospect consists of masses at up to three points, two of which must be at 0 and L. We want to show that under Assumption 1.6, the optimal prospect contains a single mass on the loss regime, with potentially a second mass at the origin. Suppose that the probability that the prospect takes a positive value and the mean of gains element of the prospect are given. Then the probability of the prospect taking a value on losses, and the mean loss z may also be considered as given.

Consider prospects on losses in form[20] of $\mathcal{P} = (L, p_L; x, p_x; 0, p_0)$ with $L < x < 0$. We have the relationships $p_L + p_x + p_0 = \phi$ and $Lp_L + xp_x = -z$ for fixed $\phi \in (0, 1)$ and $z \in (0, -L]$. To show that under the stated assumptions the optimal prospect on losses is a two-point distribution, it is sufficient to show the existence of some feasible two-point prospects which are at least as good as $\mathcal{P}$.

Recall that w_- is an inverse-S shaped function which is concave on $[0, q_-]$ and convex on $[q_-, 1]$. By the Elasticity Assumption, $E(p; w_-, c)$ is decreasing in $p \in [0, \min(c, q_-)]$ for any $c \in [0, 1]$. The prospect value of $\mathcal{P}$ is given by

$$V = w_-(p_L)v(L) + (w_-(p_L + p_x) - w_-(p_L))v(x) = w_-(p_L)v(L) + (w_-(\phi - p_0) - w_-(p_L))v(x).$$

[20] Following Barberis (2012), we write a prospect $\mathcal{P}$ corresponding to a discrete random variable with n atoms at $x_1 < x_2 < ... < x_n$ of sizes $p_1, p_2, ..., p_n$ as $\mathcal{P} = (x_1, p_1; x_2, p_2; ...; x_n, p_n)$.

Since $L \leq x \leq 0$ we must have $L(\phi - p_0) \leq -z$ else there is no feasible solution. Fix p_0 and z and consider varying x, p_L and p_x. The feasible range of x is given by $L \leqslant x \leqslant -\frac{z}{\phi - p_0}$. From the mean constraint $p_L L + p_x x = -z$ and the fact that $p_L + p_x = \phi - p_0$, we have

$$\frac{dp_L}{dx} = -\frac{dp_x}{dx} = \frac{p_x}{x - L}.$$

Differentiation of the prospect value function with respect to x gives

$$\begin{aligned}
\frac{\partial V}{\partial x} &= (w_-(\phi - p_0) - w_-(p_L))v'(x) - \frac{p_x}{x - L} w'_-(p_L)(v(x) - v(L)) \\
&= \frac{(w_-(\phi - p_0) - w_-(p_L))(v(x) - v(L))}{x - L} \left(\frac{(x - L)v'(x)}{v(x) - v(L)} - \frac{p_x w'_-(p_L)}{w_-(\phi - p_0) - w_-(p_L)} \right) \\
&= \frac{(w_-(\phi - p_0) - w_-(p_L))(v(x) - v(L))}{x - L} (E(x; v, L) - E(p_L; w_-, \phi - p_0)).
\end{aligned}$$

Case 1: $p_x + p_L = \phi - p_0 \leqslant q_-$.
Then $p_L \leq q_- \wedge (\phi - p_0)$ and $E(p_L; w_-, \phi - p_0)$ is decreasing in p_L and in turn decreasing in x. Together with the fact that $E(x; v, L)$ is increasing in x, $\frac{\partial V}{\partial x}$ is either positive for all $x \in [L, -\frac{z}{\phi - p_0}]$, negative for all x over the same range, or changes sign from negative to positive as x increases. Hence, either V is monotonic, or V has a minima, and the maximal prospect value is attained at either $x = -\frac{z}{\phi - p_0}$ or $x = L$. The corresponding prospects are $(-\frac{z}{\phi - p_0}, \phi - p_0; 0, p_0)$ and $(L, \frac{z}{|L|}; 0, \phi - \frac{z}{|L|})$. Since $z \leq (\phi - p_0)|L|$ we have $\phi - \frac{z}{|L|} \geqslant p_0 \geqslant 0$ and both prospects are feasible two-point solutions with at most one mass at a non-zero location.

Case 2: $\phi - p_0 > q_-$ *and* $p_L \leq q_-$.
Then again by the fact that $E(p_L; w_-, \phi - p_0)$ is decreasing in p_L and $E(x; v, L)$ is increasing in x, the maximal prospect value (as we let p_L range between 0 and q_-) is attained at either $p_L = 0$, whence $x = -\frac{z}{\phi - p_0}$, or $p_L = q_-$, whence $x = -\frac{z + Lq_-}{\phi - q_- - p_0}$.

The former corresponds to a feasible two-point prospect $(-\frac{z}{\phi - p_0}, \phi - p_0; 0, p_0)$. The latter corresponds to a prospect $(L, q_-; -\frac{z + Lq_-}{\phi - q_- - p_0}, \phi - p_0 - q_-; 0, p_0)$. We show in the next case that is not an optimal prospect, since the prospect value can be further increased by increasing p_L above q_-.

Case 3: $\phi - p_0 > q_-$ *and* $p_L \geq q_-$.
We compare the prospect $(L, p_L; x, p_x; 0, p_0)$ with another feasible prospect which places all its mass at L and 0 and show that the latter has at least as large a PT value as the former.

We have

$$\begin{aligned} V &= w_-(p_L)v(L) + (w_-(\phi - p_0) - w_-(p_L))v(x) \\ &\leqslant w_-(p_L)v(L) + (w_-(\phi - p_0) - w_-(p_L))\frac{|x|}{|L|}v(L) \\ &= v(L)\left(\frac{|x|}{|L|}w_-(\phi - p_0) + \left(1 - \frac{|x|}{|L|}\right)w_-(p_L)\right) \\ &\leqslant v(L)w_-\left(\frac{|x|}{|L|}(\phi - p_0) + \left(1 - \frac{|x|}{|L|}\right)p_L\right) \\ &= v(L)w_-\left(p_L + \frac{|x|}{|L|}(\phi - p_0 - p_L)\right) \end{aligned}$$

where we have used the fact that v is convex on $[L, 0]$ in the second line and the fact that w_- is convex on $[q_-, 1]$ in the fourth line. But $v(L)w\left(p_L + \frac{x}{L}(\phi - p_0 - p_L)\right)$ is the value of a feasible two-point prospect $\left(L, p_L + \frac{|x|}{|L|}(\phi - p_0 - p_L); 0, p_0 + (1 - \frac{|x|}{|L|})(\phi - p_0 - p_L)\right)$. □

We saw in Proposition 1.4 that the assumptions of Proposition 1.14 are satisfied by a large class of value functions and probability weighting functions including those commonly considered in the literature.

1.B.3 The combined problem for gains and losses

With the solutions from the previous sub-problems for gains and losses, the combined-problem is to find

$$\sup_{(\phi_\pm, z_\pm) \in H} \left(D_+(\phi_+, z_+) + D_-(\phi_-, z_-)\right)$$

where $H = \{(\phi_\pm, z_\pm) : \phi_\pm \geqslant 0, \phi_+ + \phi_- \leqslant 1, z_+ - z_- = X_0, z_+ \geqslant X_0^+, z_- \in [X_0^-, -L\phi_-], \}$. The optimal prospect can be identified on solving this problem.

Proof of Proposition 1.7. If $(\phi_\pm^*, z_\pm^*)$ are the optimisers to the combined problem, then the optimiser for (1.25) is given by combining the optimisers for the sub-problems of gains and losses such that the optimal quantile function is given by

$$G^*(x) = G_-(x; \phi_-^*, z_-^*)1_{(0,\phi_-^*]} + G_+(x; \phi_+^*, \psi^*, \lambda^*)1_{(1-\phi_+^*,1)}.$$

where (ψ^*, λ^*) are the optimisers to the sub-problem for gains with $z_+ = z_+^*$ and $\phi_+ = \phi_+^*$. Expressing z_-^* in terms of z_+^*, we can rewrite the constraint on $z_\pm^*$ as $X_0^+ \leqslant z_+^* \leqslant X_0 - L\phi_-$.

Under the assumption that $v'(0+) = \infty$, by the remarks at the end of the proof of Lemma 1.13 we have $\phi_-^* = 1 - \phi_+^*$. Moreover, suppose that a candidate optimal solution of the problem includes a mass at the origin in the loss-component. We could reclassify this mass as part of the gain-distribution. But then, the prospect value could be improved by redistributing this mass to become strict gains. Hence the candidate solution cannot be optimal, and there cannot be any mass at zero in the optimal prospect. In particular, the form of the optimal quantile function is given by (1.11).

□

In summary, the optimal prospect does not allocate any probability mass to zero (which corresponds to the reference level prior to scaling). Moreover, Assumption 1.6 provides a simple sufficient (and decoupled) condition on the behaviours of v and w_- leading to the feature that there is only one single atom on loss.

1.C Proof of the well-posedness condition under the base model of Tversky and Kahneman (1992)

Proof of Proposition 1.1. Let $\widehat{w}(p) = p^{\delta_+}$. Then obviously $\lim_{p\downarrow 0} \frac{\widehat{w}(p)}{w_+(p)} = 1$. Then for some fixed $\epsilon > 0$ there exists $p^* > 0$ such that $\widehat{w}(p) < (\epsilon+1)w_+(p)$ for $0 < p < p^*$.

Consider a two-point zero-mean prospect $(-a, \frac{b}{a+b}; b, \frac{a}{a+b})$ with b being large enough such that $\frac{a}{a+b} < p^*$. Then the prospect value is given by

$$\begin{aligned} V(a,b) &= w_+\left(\frac{a}{a+b}\right)v(b) + w_-\left(\frac{b}{a+b}\right)v(-a) \\ &\geqslant \frac{1}{\epsilon+1}\widehat{w}\left(\frac{a}{a+b}\right)v(b) + v(-a) \\ &= \frac{1}{\epsilon+1}\left(\frac{a}{a+b}\right)^{\delta_+}((b+R^\beta)^{\frac{1}{\beta}} - R)^{\alpha_+} + v(-a) \\ &\to \infty \end{aligned}$$

as $b \to \infty$ if $\delta_+ < \frac{\alpha_+}{\beta}$.

To prove the well-posedness property under $\delta_+ > \frac{\alpha_+}{\beta}$, it is sufficient to show that the gain-part value $D_+(\phi_+, \mu)$ is finite for any ϕ_+ and μ. Using the cumulative distribution function formulation (see Lemma 3.1 of Xu and Zhou (2013)), we can rewrite the gain-part value as

$$D_+(\phi_+, z) = \sup_{\overline{F}\in B_{\phi_+,z}} \int_0^\infty w_+(\overline{F}(x))v'(x)dx$$

where

$$B_{\phi,z} = \{\overline{F} |\ \overline{F} : [0,\infty) \to [0,1] \text{ is a decreasing function}, \int_0^\infty \overline{F}(x)dx = z, \overline{F}(0) = \phi\}.$$

Since $w_+(p) \leqslant \widehat{w}(p) = p^{\delta_+}$ for all p, it is sufficient to show that $\widehat{D}_+(\phi_+, z)$ is finite where $\widehat{D}_+(\phi_+, z) := \sup_{\overline{F} \in B_{\phi_+, z}} \int_0^\infty \widehat{w}(\overline{F}(x)) v'(x) dx$. Since $\widehat{w}$ is concave, the optimiser for $\widehat{D}$ can be obtained by solving a simple Lagrangian problem where the solution is

$$\begin{aligned} \overline{F}^*(x) &= \min\left(\phi_+, (\widehat{w}')^{-1}\left(\frac{\lambda}{v'(x)}\right)\right) \\ &= \min\left(\phi_+, \left(\left(\frac{\alpha_+\delta_+}{\lambda\beta}\right)\frac{1}{(x+R^\beta)^{\frac{\beta-1}{\beta}}((x+R^\beta)^{\frac{1}{\beta}} - R)^{1-\alpha_+}}\right)^{\frac{1}{1-\delta_+}}\right) \end{aligned}$$

and the optimal value is in form of

$$\begin{aligned} \widehat{D}_+(\phi_+, \mu) &= \int_0^\infty w_+(\overline{F}^*(x)) v'(x) dx \\ &= C + \int_K^\infty \left(\frac{\delta_+}{\lambda}\right)^{\frac{\delta_+}{1-\delta_+}} \left(\left(\frac{\alpha_+}{\beta}\right)\frac{1}{(x+R^\beta)^{\frac{\beta-1}{\beta}}((x+R^\beta)^{\frac{1}{\beta}} - R)^{1-\alpha_+}}\right)^{\frac{1}{1-\delta_+}} dx \end{aligned}$$

for some constants C and K. This indefinite integral is convergent if $\delta_+ > \frac{\alpha_+}{\beta}$. □

1.D On Elasticity measures of popular value and probability weighting functions

Proof of Proposition 1.3. The fact that $E(x; ax+b; c) = 1$ is immediate from the definition. Also

$$\begin{aligned} E(x; f, c)E(f(x); g, f(c)) = \frac{(x-c)f'(x)}{f(x)-f(c)}\frac{(f(x)-f(c))g'(f(x))}{g(f(x))-g(f(c))} &= \frac{(x-c)(g\circ f)'(x)}{(g\circ f)(x) - (g\circ f)(c)} \\ &= E(x; (g\circ f), c). \end{aligned}$$

□

For several popular classes of convex value functions v and inverse-S shaped probability weighting functions w, this result allows us to give simple proofs to show that $E(x; v, c)$ is increasing in $x \geqslant c$ for any c, and $E(p; w, c)$ is decreasing on $p \in [0, \min(q, c)]$ for any $c \in [0, 1]$ where q is the inflexion point of w. The remainder of this appendix is devoted to giving such proofs.

1.D.1 Value functions

Power function

Suppose v has the form of $v(x) = x^\alpha$ defined on $[0, \infty)$ and $\alpha > 0$. Then

$$E(x; v, c) = \frac{\alpha(x-c)x^{\alpha-1}}{x^\alpha - c^\alpha}$$

with $c \geqslant 0$. Differentiation gives

$$E'(x; v, c) = \frac{\alpha x^{\alpha-2}(cx^\alpha - \alpha x c^\alpha + (\alpha-1)c^{\alpha+1})}{(x^\alpha - c^\alpha)^2}.$$

Consider $H(x) = cx^\alpha - \alpha x c^\alpha + (\alpha-1)c^{\alpha+1}$ and note that $H(c) = 0$. We have $\partial H/\partial x = H'(x) = \alpha c(x^{\alpha-1} - c^{\alpha-1})$ and note that $H'(c) = 0$. Then for $\alpha > 1$ we have $H(x)$ is convex in x and $H \geq 0$. If $\alpha < 1$ then H is concave in x and $H \leq 0$. It follows that E is monotonic increasing in x if $\alpha > 1$ and monotonic decreasing in x if $\alpha < 1$.

Exponential function

For $v(x) = e^{\alpha x}$ on $[0, \infty)$ and $\alpha > 0$,

$$E(x; v, c) = \frac{\alpha(x-c)e^{\alpha x}}{e^{\alpha x} - e^{\alpha c}} = \frac{\alpha(x-c)e^{\alpha(x-c)}}{e^{\alpha(x-c)} - 1}.$$

Note that $e^{\alpha(x-c)} > 1 + \alpha(x-c)$. Then differentiation gives

$$E'(x; v, c) = \frac{\alpha e^{\alpha(x-c)}(e^{\alpha(x-c)} - \alpha(x-c) - 1)}{(e^{\alpha(x-c)} - 1)^2} \geqslant 0$$

so that $E(x; v, c)$ is monotonic increasing in x.

Reverse-power function

Consider $v(x) = K^\alpha - (K-x)^\alpha$ on $x \in [0, K]$ for some $K \geqslant 0$ and $0 < \alpha < 1$. Then v is a non-negative convex function with $v(0) = 0$. For $x, c \in [0, K]$, we have using Proposition 1.3 twice,

$$E(x; v, c) = E(x; (K-x)^\alpha, c) = E(K-x; x^\alpha, K-c).$$

Since $E(y; x^\alpha, z)$ is decreasing in y we conclude that $E(x; v, c)$ is increasing in x.

Scaled power function from TK value function

Consider $v(x) = -k(h - (x + h^\beta)^{1/\beta})^\alpha$ for $-h^\beta < x < 0$ with $\alpha, \beta > 0$. Then, for $-h^\beta \leq c < 0$,

$$\begin{aligned}
E(x; v, c) &= E(x; v(z) = (h - (z + h^\beta)^{1/\beta})^\alpha, c) \\
&= E(x + h^\beta; v(z) = (h - z^{1/\beta})^\alpha, c + h^\beta) \\
&= E((x + h^\beta)^{1/\beta}; v(z) = (h - z)^\alpha, (c + h^\beta)^{1/\beta})E(x + h^\beta; y^{1/\beta}, c + h^\beta) \\
&= E((x + h^\beta)^{1/\beta}; v(z) = h^\alpha - (h - z)^\alpha, (c + h^\beta)^{1/\beta})E(x + h^\beta; y^{1/\beta}, c + h^\beta).
\end{aligned}$$

Now suppose that $0 < \alpha, \beta < 1$. Then $0 < \alpha < 1 < 1/\beta$. From the discussion of the power and reverse-power functions, both of $E(h - (x + h^\beta)^{1/\beta}; v(z) = z^\alpha, h - (c + h^\beta)^{1/\beta})$ and $E(x + h^\beta; y^{1/\beta}, c + h^\beta)$ are increasing in x. Hence the product is increasing in x.

1.D.2 Probability weighting functions

Let w be an inverse-S shaped weighting function which is concave on $[0, q]$ and convex on $[q, 1]$. We want to show that for common families of weighting function $E(x; w, c)$ is decreasing in x for $0 \leqslant x \leqslant \min(c, q)$ for any $0 \leqslant c \leqslant 1$.

Let $G(x, c) = \frac{\partial}{\partial x} \ln E(x; w, c) = -\frac{1}{c-x} + \frac{w''(x)}{w'(x)} + \frac{w'(x)}{w(c)-w(x)}$. It is necessary and sufficient to show that $G(x, c) \leqslant 0$ for all $0 \leqslant x \leqslant \min(c, q)$ and $0 \leqslant c \leqslant 1$. Fix some x with $0 \leqslant x \leqslant q$. Then by a repeated application of l'Hôpital's rule

$$\lim_{c \downarrow x} G(x, c) = \lim_{c \downarrow x} \left(-\frac{1}{c - x} + \frac{w''(x)}{w'(x)} + \frac{w'(x)}{w(c) - w(x)} \right) = \frac{w''(x)}{2w'(x)} \leqslant 0$$

since $w''(x) \leqslant 0$ for $x \leqslant q$. Then, if one could show that $G(x, c)$ is decreasing in $c \in [x, 1]$, then $G(x, c) \leqslant G(x, x) \leqslant 0$, and the result will follow.

Hence, for our desired conclusion it is sufficient to show that $\frac{\partial}{\partial c} G(x, c) = \frac{1}{(c-x)^2} - \frac{w'(x)w'(c)}{(w(c)-w(x))^2} \leqslant 0$ for $x \leq c$, or equivalently $f(x, c) = f(x, c; w) \geq 0$ for $x \leq c$ where

$$f(x, c) = w'(c)w'(x) - \left(\frac{w(c) - w(x)}{c - x} \right)^2. \tag{1.31}$$

In the remainder of this section, we use this approach to prove the results for the Goldstein and Einhorn (1987) and Prelec (1998) weighting functions, and give some analysis for the Tversky and Kahneman (1992) function.

The Goldstein and Einhorn (1987) weighting function

The Goldstein and Einhorn (1987) weighting function is given by $w(x) = \frac{\gamma x^d}{\gamma x^d + (1-x)^d}$ with $0 < \gamma, d < 1$.

Differentiation gives

$$w'(x) = \frac{d}{\gamma x^2} \left(\frac{1}{x} - 1 \right)^{d-1} \left(1 + \frac{1}{\gamma} \left[\frac{1}{x} - 1 \right]^d \right)^{-2}$$

and in turn, with $z = \frac{1}{x} - 1$ and $y = \frac{1}{c} - 1$

$$\begin{aligned} f(x, c) &= \frac{d^2}{\gamma^2 x^2 c^2} z^{d-1} y^{d-1} \left(1 + \frac{z^d}{\gamma} \right)^{-2} \left(1 + \frac{y^d}{\gamma} \right)^{-2} - \frac{1}{(c - x)^2} \left(\frac{1}{1 + \frac{z^d}{\gamma}} - \frac{1}{1 + \frac{y^d}{\gamma}} \right)^2 \\ &= \frac{1}{\gamma^2 (c - x)^2} \left(1 + \frac{z^d}{\gamma} \right)^{-2} \left(1 + \frac{y^d}{\gamma} \right)^{-2} \left(d^2 \left(\frac{1}{x} - \frac{1}{c} \right)^2 z^{d-1} y^{d-1} - \left(z^d - y^d \right)^2 \right). \end{aligned}$$

Since $x \leq c$ we have $y \leqslant z$. The condition for $f(x, y) \geqslant 0$ is then

$$d^2(z-y)^2 z^{d-1} y^{d-1} - (z^d - y^d)^2 \geqslant 0. \tag{1.32}$$

On writing $z = \lambda y$ with $\lambda \geqslant 1$, (1.32) is equivalent to $d(\lambda - 1)\lambda^{\frac{d}{2}-\frac{1}{2}} - (\lambda^d - 1) \geqslant 0$ and in turn $g_d(\lambda) \geq 0$ where

$$g_d(\lambda) := d(\lambda^{1/2} - \lambda^{-1/2}) - (\lambda^{d/2} - \lambda^{-d/2}) \geqslant 0. \tag{1.33}$$

Write $\lambda = e^{2\theta}$ for $\theta \geq 0$. Then $g_d(\lambda) = h(\theta)$ where $h(\theta) = 2d \sinh\theta - 2\sinh(d\theta)$. But $h(0) = 0$ and $h'(\theta) = 2d[\cosh\theta - \cosh(d\theta)] \geq 0$ since $d < 1$. Hence g_d is increasing in λ for $\lambda \geq 1$. Since $g_d(1) = 0$ the result follows.

The Prelec (1998) weighting function

The Prelec (1998) probability weighting function is given by $w(x) = \exp(-b(-\ln x)^a)$ for $0 \leqslant x \leqslant 1$ where $0 < a < 1$ and $0 < b \leq 1$. It is sufficient to show that $f(x, c; w) \geq 0$ where f is given by (1.31). For the Prelec weighting function, we have

$$w'(x) = \frac{ab}{x}(-\ln x)^{a-1} \exp(-b(-\ln x)^a).$$

Let $y = -\ln x > 0$ and $z = -\ln c > 0$, then $y \geqslant z \geqslant 0$ and $f(x, c) = H(y, z)$ where

$$\begin{aligned}
H(y,z) &:= w'(e^{-z})w'(e^{-y}) - \left(\frac{w(e^{-z}) - w(e^{-y})}{e^{-z} - e^{-y}}\right)^2 \\
&= (ab)^2 (yz)^{a-1} e^{y+z-b(y^a+z^a)} - \left(\frac{e^{-bz^a} - e^{-by^a}}{e^{-z} - e^{-y}}\right)^2 \\
&= \frac{e^{-b(y^a+z^a)}}{(e^{-z} - e^{-y})^2}\left((ab)^2(yz)^{a-1}\left(e^{\frac{1}{2}(y-z)} - e^{-\frac{1}{2}(y-z)}\right)^2 - \left(e^{\frac{b}{2}(y^a-z^a)} - e^{-\frac{b}{2}(y^a-z^a)}\right)^2\right) \\
&= \frac{4e^{-b(y^a+z^a)}}{(e^{-z} - e^{-y})^2}\left((ab)^2(yz)^{a-1}\sinh^2\left(\frac{1}{2}(y-z)\right) - \sinh^2\left(\frac{b}{2}(y^a - z^a)\right)\right).
\end{aligned}$$

Hence to show $H(y, z) \geqslant 0$, it is equivalent to show that

$$(ab)^2(yz)^{a-1}\sinh^2\left(\frac{1}{2}(y-z)\right) - \sinh^2\left(\frac{b}{2}(y^a - z^a)\right) \geqslant 0$$

or

$$\frac{\sinh\left(\frac{1}{2}(y-z)\right)}{\sinh\left(\frac{b}{2}(y^a - z^a)\right)} \geqslant \frac{(yz)^{\frac{1}{2}(1-a)}}{ab}. \tag{1.34}$$

Let $\gamma = \frac{z}{y}$ such that $0 \leqslant \gamma \leqslant 1$. Then the condition in (1.34) becomes

$$\frac{\sinh\left(\frac{y}{2}(1-\gamma)\right)}{\sinh\left(\frac{by^a}{2}(1-\gamma^a)\right)} \geqslant \frac{y^{1-a}\gamma^{\frac{1}{2}(1-a)}}{ab}. \tag{1.35}$$

It is known that under $b \leqslant 1$, the inflexion point of w lies above the diagonal line. Thus q is smaller than the fixed point of w which is given by $\exp(-b^{\frac{1}{1-a}})$, i.e. $q \leqslant \exp(-b^{\frac{1}{1-a}})$. Then $-y = \ln x \leqslant \ln q \leqslant -b^{\frac{1}{1-a}}$ and in turn $y \geqslant by^a$. Also we have $1-\gamma \geqslant 1-\gamma^a$, then $\frac{y}{2}(1-\gamma) \geqslant \frac{by^a}{2}(1-\gamma^a)$. By convexity of the function $\sinh(\cdot)$ on the positive regime,

$$\frac{\sinh\left(\frac{y}{2}(1-\gamma)\right)}{\sinh\left(\frac{by^a}{2}(1-\gamma^a)\right)} \geqslant \frac{\frac{y}{2}(1-\gamma)}{\frac{by^a}{2}(1-\gamma^a)} = \frac{y^{1-a}(1-\gamma)}{b(1-\gamma^a)}.$$

Hence to show (1.35) it is sufficient to show

$$\frac{y^{1-a}(1-\gamma)}{b(1-\gamma^a)} \geqslant \frac{y^{1-a}\gamma^{\frac{1}{2}(1-a)}}{ab}$$

or equivalently

$$G(\gamma) := a(1-\gamma) - \gamma^{\frac{1}{2}(1-a)}(1-\gamma^a) \geqslant 0$$

on $0 \leqslant \gamma \leqslant 1$. But $\gamma^{-1/2}G(\gamma) = a(\gamma^{-1/2} - \gamma^{1/2}) - (\gamma^{-a/2} - \gamma^{a/2}) = g_a(1/\gamma)$ where g_a is the function given in (1.33) and shown to be positive there.

The Tversky and Kahneman (1992) weighting function

The Tversky and Kahneman (1992) probability weighting function is given in (1.3) as: $w(x) = \frac{x^\delta}{(x^\delta+(1-x)^\delta)^{1/\delta}}$. The decreasing elasticity property on the concave regime seems difficult to verify analytically. In Figure 1.7 we plot the $-G(x,c) = -\frac{\partial}{\partial x}\ln E(x;w,c) = \frac{1}{c-x} - \frac{w''(x)}{w'(x)} - \frac{w'(x)}{w(c)-w(x)}$ for several values of δ over $0 \leqslant x \leqslant \min(c,q)$ and $0 \leqslant c \leqslant 1$ where q is the inflexion point of w. All the plots show positive values which verify the decreasing elasticity property of w on the required range.

1.E Proofs for the implied rate of selling

Our goal in this section is to prove Lemma 1.10 which can be established by the following collection of results. To begin with, the key mathematical result which underpins our analysis is:

Lemma 1.15 (Bertoin et al. (1992))**.** *Let X be a diffusion process in natural scale with $X_0 = 0$ and satisfying $dX_t = \xi(X_t)dB_t$. Suppose that ν is a centred target law. Let $\zeta = \zeta_\nu$ be given by $\zeta(dx) = \frac{\nu(dx)}{u_\nu(x)-|x|}$ and let $\Lambda_t^{X,\zeta} = \int_{\mathbb{R}} \zeta(dx) L_t^X(x)$ where $L^X = (L_t^X(x))_{t\geq 0, s(a_J)\leq x\leq s(b_J)}$ is the local time process of X defined in the same way as in the proof of Proposition 1.8. Set $\tau = \tau^{X,\zeta,\Lambda} = \inf\{u : \Lambda_u^{X,\zeta} \geq T\}$ where T is a standard exponential random variable with unit rate which is independent of X. Then $X_\tau \sim \nu$.*

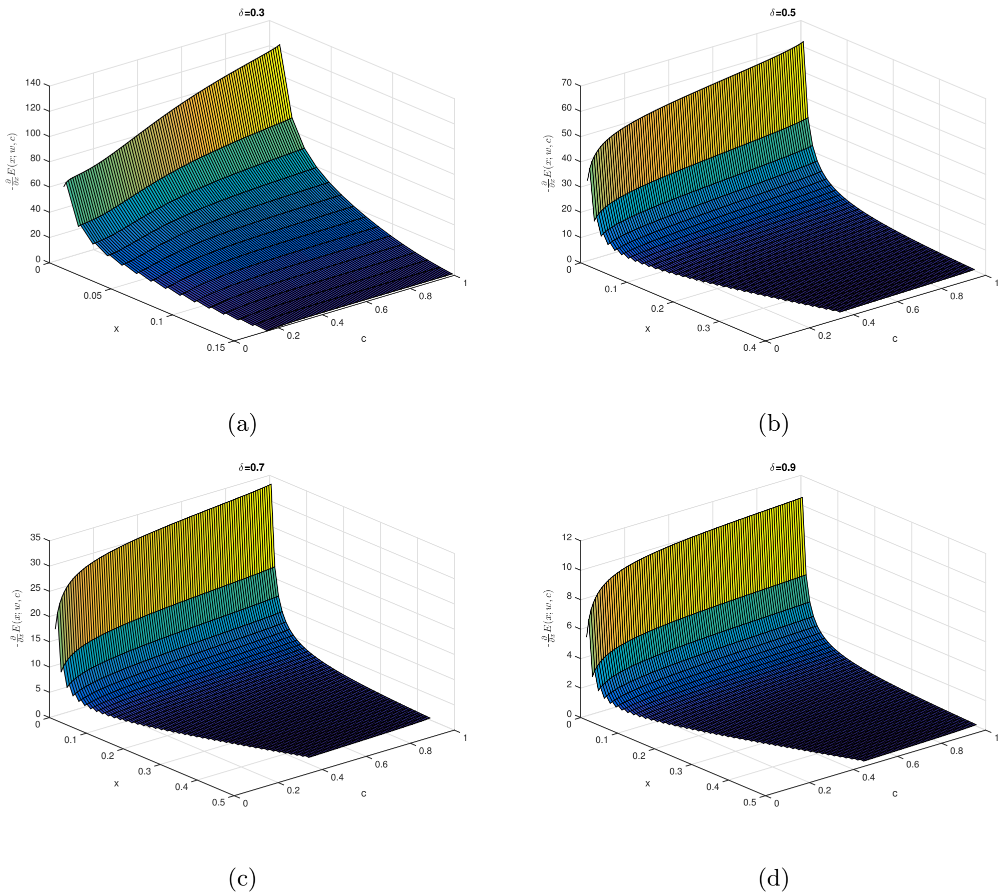


Figure 1.7: Plot of $-G(x, c) = -\frac{\partial}{\partial x} \ln E(x; w, c)$ with w taken to be the Tversky and Kahneman (1992) probability weighting function for several values of parameter δ. To conclude that $E(x; w, c)$ is decreasing in x for $0 \leq x \leq \min\{c, q\}$ we need $-G(x, c) \geq 0$ over the relevant range.

For our purposes it is convenient to work with a slightly different normalisation of local time which satisfies an occupation time formula for calendar time, rather than quadratic variation. Set $\ell_t^X(a) = L_t^X(a)/\xi(a)^2$. Then, for a Borel function Φ,

$$\int_0^t \Phi(X_u)\xi(X_u)^2 du = \int_0^t \Phi(X_u) d[X]_u = \int L_t^X(a)\Phi(a)da = \int \ell_t^X(a)\Phi(a)\xi(a)^2 da$$

and we have $\int_0^t \Psi(X_u)du = \int \ell_t^X(a)\Psi(a)da$. Set $\Gamma_t^{X,\eta} = \int \eta(dx)\ell_t^X(x)$. Then $\Lambda_t^{X,\eta} = \Gamma_t^{X,\xi^2\eta}$ where $\xi^2\eta$ is the measure $\xi^2(x)\eta(dx)$.

Proposition 1.16. *Let $\lambda = \lambda_\nu$ be given by $\lambda(dx) = \frac{\xi(x)^2\nu(dx)}{u_\nu(x)-|x|}$. Let $\tau = \tau^{X,\lambda,\Gamma} = \inf\{u : \int \lambda(da)\ell_t^X(a) \geq T\}$. Then $X_\tau \sim \nu$.*

Proof. This directly follows from Lemma 1.15 and construction of λ. $\square$

Let ℓ^P be the local time process of P with the normalisation that $\int_0^t \Phi(P_s)ds = \int \Phi(a)\ell_t^P(a)da$.

Proposition 1.17. *Suppose P is given as the solution to the SDE $dP_t = \sigma(P_t)dB_t + \kappa(P_t)dt$. Let s be a scale function of P chosen such that $s(P_0) = 0$. Let μ be a prospect such that $\nu := \mu \circ s$ has zero mean.*

Let $\gamma = \gamma_\mu$ be given by $\gamma(dp) = \sigma(p)^2 s'(p)\frac{\mu(dp)}{u_\nu(s(p))-|s(p)|}$ and let $\Gamma_t^{P,\gamma} = \int_{\mathbb{R}} \gamma(dp)\ell_t^P(p)$. Set $\tau = \tau^{P,\gamma,\Gamma} = \inf\{u : \Gamma_u^{P,\gamma} \geq T\}$ where T is a standard exponential random variable with unit rate which is independent of P. Then $P_\tau \sim \mu$.

Proof. Let $X = s(P)$. For a Borel function Ψ we have

$$\begin{aligned}\int \Psi(s(p))\ell_t^X(s(p))s'(p)dp = \int \Psi(x)\ell_t^X(x)dx &= \int_0^t \Psi(X_u)du \\ &= \int_0^t (\Psi\circ s)(P_u)du = \int (\Psi \circ s)(p)\ell_t^P(p)dp.\end{aligned}$$

In particular, $\ell_t^X(s(p))s'(p) = \ell_t^P(p)$.

From Itô's formula we have $dX_u = \xi(X_u)dB_u = \sigma(P_u)s'(P_u)dB_u$ and so $\xi(s(p)) = s'(p)\sigma(p)$. Then

$$\begin{aligned}\Gamma_u^{P,\gamma} &= \int \frac{\mu(dp)}{u_\nu(s(p))-|s(p)|}\sigma(p)^2 s'(p)\ell_u^P(p) = \int \frac{\mu(dp)}{u_\nu(s(p))-|s(p)|}\sigma(p)^2[s'(p)]^2\ell_u^X(s(p)) \\ &= \int \frac{\mu(dp)}{u_\nu(s(p))-|s(p)|}\xi(s(p))^2\ell_u^X(s(p)) = \int \frac{\nu(dx)}{u_\nu(x)-|x|}\xi(x)^2\ell_u^X(x) = \Gamma_u^{X,\lambda}\end{aligned}$$

where λ is as given in Proposition 1.16. Then $\tau = \tau^{P,\gamma,\Gamma} = \tau^{X,\lambda,\Gamma}$ and $s(P_\tau) = X_\tau = X_{\tau^{X,\lambda,\Gamma}} \sim \nu$. We conclude that $P_\tau \sim \mu$. $\square$

1.F Distribution of agents' types and aggregate implied selling rate

In this section we discuss how to design the distribution of agents' types to produce a flat aggregate implied selling rate on losses as in Figure 1.6. We assume the agents in the economy have homogeneous TK preference parameters except their loss aversion levels k vary on $[k_{\min}, \infty)$. Here $k_{\min} := \left(\frac{R-\ell_1}{R}\right)^{\alpha_- - \alpha_+}$ with R being the common reference level and ℓ_1 the target loss threshold of the agent with $k = 1$. The price process is a martingale exponential Brownian motion starting at the reference level such that $\beta = 1$ and $P_0 = R$. We insist $\alpha_- > \alpha_+$.

Lemma 1.18 (A scaling property)**.** *Let F_k be the CDF of the optimal prospect of the agent with loss aversion level k. Then $F_k(p) = F_1(R + k^{\frac{1}{\alpha_- - \alpha_+}}(p - R))$ for $k \geqslant k_{\min}$. Moreover, all agents allocate a common size θ_- of probability mass to losses, and the target loss threshold of the agent with loss aversion level k is given by $\ell_k = R - k^{-\frac{1}{\alpha_- - \alpha_+}}(R - \ell_1)$.*

Proof. Under the assumption of $\beta = 1$, the scaled value function of the agent with loss aversion level k is simply given by

$$v_k(x) = \begin{cases} x^{\alpha_+}, & x > 0, \\ -k|x|^{\alpha_-}, & x \leqslant 0, \end{cases}$$

and his corresponding PT objective function is

$$\mathcal{E}_k(X) = \int_{1-\mathbb{P}(X>0)}^{1} v_k(G(x))w'_+(1-x)dx + \int_0^{\mathbb{P}(X<0)} v_k(G(x))w'_-(x)dx$$

with X belonging to a class of random variables with zero mean and lower bound $-R$, and G is the quantile function of X. Then we have

$$\begin{aligned}
\mathcal{E}_k(k^{-\frac{1}{\alpha_- - \alpha_+}} X) &= \int_{1-\mathbb{P}(k^{-\frac{1}{\alpha_- - \alpha_+}} X>0)}^{1} v_k(k^{-\frac{1}{\alpha_- - \alpha_+}} G(x))w'_+(1-x)dx \\
&\quad + \int_0^{\mathbb{P}(k^{-\frac{1}{\alpha_- - \alpha_+}} X<0)} v_k(k^{-\frac{1}{\alpha_- - \alpha_+}} G(x))w'_-(x)dx \\
&= k^{-\frac{\alpha_+}{\alpha_- - \alpha_+}} \int_{1-\mathbb{P}(X>0)}^{1} v_1(G(x))w'_+(1-x)dx \\
&\quad + k^{1-\frac{\alpha_-}{\alpha_- - \alpha_+}} \int_0^{\mathbb{P}(X<0)} v_1(G(x))w'_-(x)dx \\
&= k^{-\frac{\alpha_+}{\alpha_- - \alpha_+}} \mathcal{E}_1(X).
\end{aligned}$$

Hence if X_1^* is an optimiser to the problem $\sup_X \mathcal{E}_1(X)$, then $X_k^* := k^{-\frac{1}{\alpha_- - \alpha_+}} X_1^*$ is the optimiser to the problem $\sup_X \mathcal{E}_k(X)$ (the zero-mean property trivially preserves, and the lower bound constraint is guaranteed to be satisfied as long as $k \geqslant k_{\min}$). The result immediately follows. $\square$

Assume the loss aversion levels of the agents are randomly distributed with a prior density η supported on $[k_{\min}, \infty)$. The CDF of the aggregate target sale price is given by $F_{\overline{\mu}}(p) = \int_{k_{\min}}^{\infty} F_k(p)\eta(k)dk$ where F_k is the CDF of X_k^*. Note that we have $F_{\overline{\mu}}(R) = \theta_-$.

Let $f_{\overline{\mu}}$ be the density function of $F_{\overline{\mu}}$. We require a constant stopping rate λ on the loss regime $p < R = P_0$, then using (1.18) we have $f_{\overline{\mu}}(p) = \frac{2\lambda}{\sigma^2 p^2} \int_0^p (p-u) f_{\overline{\mu}}(u)du$ and in turn $F_{\overline{\mu}}$ is a solution to the ODE

$$p^2 F''(p) + 2pF'(p) - \frac{2\lambda}{\sigma^2} F(p) = 0.$$

Using the boundary conditions of $F_{\overline{\mu}}(0) = 0$ and $F_{\overline{\mu}}(R) = \theta_-$, we conclude $F_{\overline{\mu}}(p) = \theta_- \left(\frac{p}{R}\right)^{C_\lambda}$ for $p \leqslant R = P_0$, where $C_\lambda := \frac{-1+\sqrt{1+8\lambda/\sigma^2}}{2}$.

Hence to produce this target CDF on losses, we require

$$\begin{aligned}
\theta_- \left(\frac{p}{R}\right)^{C_\lambda} = \int_{k_{\min}}^{\infty} F_k(p)\eta(k)dk &= \theta_- \int_{k_{\min}}^{\infty} 1_{(p \geqslant \ell_k)} \eta(k)dk \\
&= \theta_- \int_{k_{\min}}^{\infty} 1_{\left(k \leqslant \left(\frac{R-\ell_1}{R-p}\right)^{\alpha_- - \alpha_+}\right)} \eta(k)dk \\
&= \theta_- \int_{k_{\min}}^{\left(\frac{R-\ell_1}{R-p}\right)^{\alpha_- - \alpha_+}} \eta(k)dk.
\end{aligned}$$

Differentiating both side with respect to p gives

$$\frac{C_\lambda}{R}\left(\frac{p}{R}\right)^{C_\lambda - 1} = \frac{(\alpha_- - \alpha_+)(R - l_1)^{\alpha_- - \alpha_+}}{(R-p)^{1+\alpha_- - \alpha_+}} \eta\left(\left(\frac{R-\ell_1}{R-p}\right)^{\alpha_- - \alpha_+}\right)$$

and this gives the required form of η as

$$\eta(k) = \frac{C_\lambda (R-\ell_1)}{R(\alpha_- - \alpha_+)} \left(\frac{R - (R-\ell_1)k^{-\frac{1}{\alpha_- - \alpha_+}}}{R}\right)^{C_\lambda - 1} k^{-\left(1+\frac{1}{\alpha_- - \alpha_+}\right)}$$

defined for $k \geqslant k_{\min} = \left(\frac{R-\ell_1}{R}\right)^{\alpha_- - \alpha_+}$.

Chapter 2

Randomised strategies and prospect theory in a dynamic context

"From here on in, I've decided to make all trivial decisions with a throw of the dice, thus freeing up my mind to do what it does best: enlighten and amaze."
— Sheldon Cooper, *The Big Bang Theory (Season 5, Episode 4)*

2.1 Introduction

In the asset liquidation model under prospect theory (PT) preferences introduced in Chapter 1, the agent solves for an optimal prospect at time zero and adopts a trading strategy accordingly which attains this target prospect. Implicitly, it is assumed that the agent is a precommitting one in the sense that he is able to follow the strategy devised at time zero consistently throughout the rest of the game. This indeed is just one of the several possible notions of optimality. The main subtlety of predicting an agent's behaviours in a dynamic PT model is that its probability weighting component may - and typically will - induce time inconsistency.[1]

In presence of time inconsistency, how an individual acts in a state of the world may differ from how he planned to act in that state. It becomes crucial to understand how

[1]Other types of economic problems featuring time inconsistency include dynamic mean-variance portfolio optimisation and intertemporal choices with hyperbolic discounting.

to deal with the time inconsistency - see Machina (1989) - is the agent naive (and thus unaware of the inconsistency), or is he sophisticated and aware of the inconsistency? And if he is sophisticated, can he find a way to commit to his initial plan, or must he acknowledge the fact that his future self will re-optimise, and potentially change strategy? In his recent discussion paper on the psychology of tail events, Barberis (2013) raises the challenge of how to best address the time inconsistency induced by probability weighting.

In a model of casino gambling in a multi-period binomial tree setting, Barberis (2012) observes and investigates the phenomenon of time inconsistency with PT agents. Three types of agent are considered: an agent with a precommitment mechanism (for short, a precommitting agent), a naive agent and a sophisticated agent.[2] The optimal behaviours of the three types of agent are solved by exhaustive numerical searches. Barberis (2012) finds PT can explain observed patterns of gambling behaviour. Each type of agent will, for certain values of PT preference parameters, gamble in a casino. The first contribution of Barberis (2012) is to show that PT provides a viable model of gambling behaviour. His second major contribution is to make a striking observation concerning the behaviour of the naive agent - that sometimes the naive agent follows a strategy which is the exact opposite of the one he planned to follow at the outset.[3]

However, the realistic predictions brought by the discrete, finite horizon model are not robust to change in the modelling setup. Ebert and Strack (2015) consider a continuous time, continuous space and infinite horizon analog of the Barberis (2012) model in which wealth from gambling follows a Brownian motion. They show that for a wide range of PT specifications (including essentially all versions which have received empirical support) the agent with commitment would prefer a stopping rule based on the first exit time of the

[2]The precommitting agent decides his strategy today and has a mechanism for making his future self follow this strategy; a naive agent is not aware of his time inconsistent preference structure, and he re-evaluates all possible strategies and takes a new decision about whether to gamble further or to stop at each point of time; the sophisticated agent is aware of his time inconsistency but is unable to commit his future self to following any given strategy. See Section 2.3.1 for a more detailed description.

[3]Suppose parameters are such that the naive agent wants to gamble at time zero. On first entering the casino, the naive agent typically plans to continue to gamble if his first bet is a success, but to stop if his first bet is unsuccessful. This is a “loss-exit” strategy. But what happens in practice? If the first gamble results in a win, then the agent re-evaluates his situation and may now find it optimal to stop. In particular, the extreme outcome of a winning gamble in every time period up to the final horizon is now less extreme, and thus the impact of probability weighting is less significant. Then the fact that the agent is risk averse (on gains) means that he stops gambling. Conversely, if the first gamble results in a loss, the agent re-evaluates, and again the probability weighting is less significant, resulting in the agent electing to continue gambling (as PT preferences are risk seeking on losses).

wealth process from a well-chosen interval to one of stopping immediately. From this fact Ebert and Strack (2015) make a drastic inference: the naive PT agent never stops. This extreme prediction of a naive agent's behaviours arguably casts doubt over the applicability of PT in a continuous dynamic model.

The extant prospect theory literature has focused on pure strategies, both in static and dynamic settings. Results on mixed (or randomised) strategies are very limited. One result is given in Wakker (1994) (see also Theorem 7.4.1 of Wakker (2010)), which considers mixed strategies in the context of a static choice of prospect in a rank dependent utility model, and relates the benefit of using mixed strategies to the concavity of the probability weighting function. There is also mention of the desire of PT agents for randomisation in the mathematical finance literature, for example, Carassus and Rasonyi (2015). In contrast, randomisation has been discussed in the ambiguity literature since the seminal work of Raiffa (1961) and is standard in other branches of economics such as game theory. In this chapter we consider the role of randomisation in the agents' set of potential strategies and re-examine the impact of probability weighting in a dynamic PT model, in both the discrete setup of Barberis (2012) and the continuous setup of Ebert and Strack (2015).

The first contribution of this chapter is to highlight that a PT agent facing a dynamic investment/stopping problem will typically benefit from following a randomised strategy.[4] We demonstrate this result in the context of the casino gambling model of Barberis (2012). In this setting the wider choice of available strategies causes the agent to gamble in a larger regime of parameters.

Our second contribution is to reconsider the continuous time model of Ebert and Strack (2015). They show how to choose an interval $[a, b]$ containing the initial wealth x such that stopping at the first exit time of the wealth process from this interval is strictly preferred to stopping immediately. Their result is true under a mild condition described as probability weighting being stronger than loss aversion, which they demonstrate holds for the most popular weighting functions. We show that the randomised strategy of sometimes stopping immediately and sometimes stopping on the first exit from $[a, b]$ is a better prospect than simply stopping on the first exit from the interval. Hence the analysis of Ebert and Strack (2015) is no longer sufficient to conclude that the naive agent never stops, once randomisation is allowed. Our result requires a condition that can be interpreted as follows: probability

[4] A similar observation is made in He et al. (2016). The main focus of He et al. (2016) is on the agent who can precommit in a discrete time, infinite horizon model. In contrast, our focus is on the implications for the continuous time setting, and especially the impact of time inconsistency and the question of whether naive agents always use a trivial strategy in this setting.

weighting on losses for nearly certain events is stronger than probability weighting on gains for rare events. We show this condition follows from the prevailing experimental finding that $w_{\pm}(1/2) < 1/2$ and demonstrate (for the same parameters as Ebert and Strack (2015)) that it holds for several popular weighting functions.

Ebert and Strack (2015) extend their arguments based on intervals to prove that for naive agents restricted to pure strategies, the optimal strategy is to always continue and gamble *"until the bitter end"*. In contrast, we show that a naive agent with a randomisation device (the ability to generate a continuous random variable), and with the ability to commit his current self to following strategies which depend on the realisation of this device, does not necessarily follow a never-stopping strategy. However, our argument needs more than a result based on intervals. One way to conclude that the optimal strategy for agents who can randomise is to always continue is to determine the optimal prospect and to show that this prospect does not carry any mass at the starting wealth level. We show that this is not true in general, and that the never-stopping result no longer holds when agents can randomise. To demonstrate this, in Section 2.5 we give a pair of examples for which we calculate the optimal prospect of an agent and show that if the initial value of the Brownian motion is at the reference level, then the optimal prospect includes an atom at the reference level. A naive agent who can randomise may stop at the reference level. The analysis is further extended in Section 2.6 to cover a more general modelling setup where we show that a naive agent may stop if his net wealth is sufficiently high. Thus the drastic conclusion of Ebert and Strack (2015) that naive PT agents never stop is no longer the unique prediction if we allow for a wider class of strategies. Naive PT agents who can follow randomised strategies may voluntarily stop gambling.

The rest of this chapter is organised as follows. In Section 2.2 we briefly describe the structure of the optimal stopping problem under PT preferences and the idea of randomisation. Section 2.3 presents the discrete model and we give numerical results confirming that randomisation can lead to improvement in economic value to different types of agent. The continuous counterpart is reviewed in Section 2.4. We first demonstrate that the pure strategy in form of a first exit rule of Ebert and Strack (2015) can be improved by mixing the strategy with sometimes stopping immediately. Then through solving two simple stylised examples in Section 2.5 as well as a model under more general specifications in Section 2.6, we show that the optimal behaviours of a naive agent may include voluntary cessation of gambling. Finally we conclude in Section 2.7.

2.2 Prospect theory and optimal stopping

2.2.1 Prospect theory preferences

We work under the same PT preferences framework as in Section 1.2.1 of Chapter 1 which we briefly recap as follows. The agent's utility or value function defined over gains and losses is U such that U is continuous, $U(0) = 0$, U is concave and increasing on $\mathbb{R}_+$ and convex and increasing on $\mathbb{R}_-$ and U satisfies the (simple) loss aversion property $U(y) + U(-y) < 0$ for all y.[5] For the probability weighting functions, we assume that $w_\pm$ is concave on $[0, Q_\pm]$ and convex on $[Q_\pm, 1]$ for some $Q_\pm \in (0, 1)$ with $w_\pm(1/2) < 1/2$. For an exogenously fixed reference level R, the PT value of a random payoff Z is given by $\mathcal{E}(Z)$ as defined in (1.2).

2.2.2 Optimal stopping

We assume that the agent's wealth from gambling or trading of assets follows a stochastic process X. For this chapter, we consider X to be either a simple random walk (Section 2.3), a Brownian motion (Section 2.4 and 2.5), or some other general time-homogeneous diffusion (Section 2.6). The strategies available to the agent correspond to stopping times τ representing when to sell an asset or stop gambling. Then the stopped value $X = X_\tau$ represents the prospect of the agent. The goal of the agent is to find

$$\sup_{\tau \in \Lambda} \mathcal{E}(X_\tau)$$

where $\mathcal{E}$ is as given in (1.2). Λ is a suitable set of stopping times depending on the context of the modelling framework.

2.2.3 Randomised strategies: a preliminary discussion

In classical optimal stopping problems in the expected utility framework, the agent faces a choice between stopping and continuing at each instant of time. If there is no probability weighting then there is no incentive to randomise: the expected payoff is linear in the probabilities and if the value from stopping is V^S and the value from continuing is V^C, then the payoff $V(\theta)$ from a randomised strategy involving continuing with probability θ is $V(\theta) = \theta V^C + (1 - \theta) V^S \leq \max\{V^S, V^C\}$. Hence there is an optimal strategy which is pure, and the ability to randomise brings no benefit to the agent. This result extends to any

[5] We do not explicitly require simple loss aversion in Chapter 1 but this condition will be useful in this chapter. For the Kahneman and Tversky (1979) value function of (1.1), it satisfies simple loss aversion if $\alpha_+ = \alpha_-$ and $k > 1$.

situation where there is no probability weighting.[6] However, if probabilities are re-weighted then typically $V(\theta) \neq \theta V^C + (1-\theta)V^S$ and the agent may benefit from mixed strategies. In fact, although our focus is on PT preferences, similar arguments will apply in any dynamic setting where probability weighting is included, for example, the rank dependent utility models of Quiggin (1982) and Yaari (1987).

Extensive experimental research documents widespread patterns of stochastic or random choice whereby subjects faced with the same decision problem on multiple occasions make different choices.[7] Agranov and Ortoleva (2015) (also Dwenger et al. (2013)) design an experiment to distinguish between different explanations for the source of stochastic choice and document evidence in support of deliberate randomisation.[8] This evidence fits with our theoretical model where it can be optimal for PT agents to randomise.

2.3 The discrete model

2.3.1 Wealth dynamics and types of the agent

An agent is offered a series of independent and identical fair gambles in which he could win or lose a unit amount. At the beginning of each period, the agent decides whether to enter the gamble (continue) or to leave (stop). Upon stopping or the end of the T^{th} period, whichever comes first, the whole game ends and the agent takes the current cumulative profit or loss as his final payoff. The evolution of the agent's net wealth can be represented by a path through a T-period recombining binomial tree, and the wealth process X follows a simple random walk. At each time point $t = 0, 1, ..., T$, his net wealth can possibly take $t + 1$ distinct values labeled by $S(1, t) > S(2, t) > ... > S(t + 1, t)$. If the initial wealth of the agent is zero then $S(i, t) = t - 2(i - 1)$. In this setup, the available strategies Λ is a set of stopping times taking values on $\{0, 1, ..., T - 1\}$.

Barberis (2012) assumes that the agent's decision at a particular node is a binary choice of exit or continuation which only depends on the current position of that node. A plan at node (i, t) is defined as a mapping $C : (i, t) \to \{stop, continue\}$. When the agent is

[6]Kyle et al. (2006) and Henderson (2012) consider optimal stopping problems for PT agents without probability weighting and determine the optimal stopping rules. Since there is no probability weighting in these settings their proposed strategies are optimal, even if randomised strategies are allowed.

[7]This pattern of stochastic choice was first noted by Tversky (1969) and replicated in many studies, see Agranov and Ortoleva (2015) and Dwenger et al. (2013) and references therein.

[8]Other theoretical models where randomisation is a deliberate choice include Machina (1985), Cerreia-Vioglio et al. (2013), Cerreia-Vioglio et al. (2015) and Saito (2015).

at node (i,t), a collection of plans at all reachable subsequent nodes $a = \{C(k,s) : t \leqslant s \leqslant T-1, i \leqslant k \leqslant s-t+i\}$ is a stopping strategy of the game associated with (i,t) as a starting position. Let $A(i,t)$ be the set of all possible stopping strategies available to the agent who is at the node (i,t). Then under a given stopping strategy $a \in A(i,t)$ which leads to a stopping time $\tau(a)$, one can compute the probability distribution of $X^{(i,t)}_{\tau(a)}$ (the superscript (i,t) is used to indicate that the wealth process starts at the node (i,t)). Assuming a zero reference level, the PT value of this strategy is computed by $\mathcal{E}(X^{(i,t)}_{\tau(a)})$. The exact behaviours of the agent depends on his own type as described below.

The precommitting agent

At time zero, the precommitting agent solves the optimisation problem

$$\sup_{a \in A(1,0)} \mathcal{E}(X^{(1,0)}_{\tau(a)}).$$

Define the corresponding optimiser as $a^* = \{C^*(i,t) : 0 \leqslant t \leqslant T-1, 1 \leqslant i \leqslant t+1\}$. The precommitting agent acts according to the plan $C^*(i,t)$ in any subsequent node (i,t). In this case, the optimal strategy is computed only once at time zero and it characterises the subsequent behaviours of the agent completely.

The naive agent

Unlike the precommitting agent, the naive agent re-computes the optimal strategy at every node and decides whether to continue or stop in the current node based on the new optimisation result. The updated decision overrides the plans derived in any previous time step. Mathematically, the agent who is currently at node (i,t) solves

$$\sup_{a \in A(i,t)} \mathcal{E}(X^{(i,t)}_{\tau(a)}).$$

Suppose the optimal solution associated with this particular starting node is given by $a^*_{(i,t)} = \{C^*_{(i,t)}(k,s) : t \leqslant s \leqslant T-1, i \leqslant k \leqslant s-t+i\}$. Then the agent's action at this node is given by $C^*_{(i,t)}(i,t)$. This type of agent is time inconsistent as his planned action at a fixed node $(\overline{K}, \overline{S})$ may change with his current position (i,t) (that is, $C^*_{(i,t)}(\overline{K}, \overline{S})$ depends on (i,t)).

The sophisticated agent

This type of agent is aware that he will re-evaluate the prospect value in future states and follow the best strategy at that moment in time. Nonetheless, he does not have any commitment device to force his future self to follow a strategy planned today. Instead, he acts in an optimal way today in anticipation of how he will behave in the future. Consequently, this type of agent operates on the logic of backward induction.[9] Starting from the last period of the game, the agent solves for the optimal plan at the nodes of $(1, T-1), (2, T-1), ..., (T, T-1)$. Knowing his behaviours in the future time, he iterates one time-step backward and decides whether it is optimal to continue or quit the game at the preceding nodes. The process is repeated all the way back to time zero. Precisely, the optimal decision C^* at each node is given by the algorithm

$$C^*(i, T-1) = \underset{C(i,T-1)\in\{\text{stop},\text{continue}\}}{\text{argmax}} \mathcal{E}(X^{(i,T-1)}_{\tau(C(i,T-1))}), \quad 1 \leqslant i \leqslant T,$$
$$C^*(i, t) = \underset{C(i,t)\in\{\text{stop},\text{continue}\}}{\text{argmax}} \mathcal{E}(X^{(i,t)}_{\tau(C(i,t)\cup a^*(i,t))}), \quad 1 \leqslant i \leqslant t+1, t < T-1,$$

where $a^*(i,t) := \{C^*(k,s) : t < s \leqslant T-1, i \leqslant k \leqslant s-t+i\}$ represents the collection of the optimal actions at all reachable nodes beyond (i,t).

2.3.2 Randomisation in the discrete model

Rather than limiting the plan at each node to a binary choice between stopping or continuing, the agent could consider randomising the stopping decision such that the probability of continuing at node (i,t) is given by $\theta(i,t)$. A stopping strategy associated with a particular starting node (i,t) is no longer a collection of binary mappings but instead a collection of continuation probabilities $\{\theta(k,s) \in [0,1] : t \leqslant s \leqslant T-1, i \leqslant k \leqslant s-t+i\}$. In this setting, the PT value of a stopping strategy can be obtained by (1.2) as before upon evaluating the payoff distribution, and one can still distinguish the three types of agent as in Barberis (2012) by amending the definitions in Section 2.3.1 to allow for a supremum over a suitable space Θ. Note that if we restrict the continuation probabilities to take values on $\{0,1\}$ only, we recover the setup of Barberis (2012).

2.3.3 Numerical results in a two-period model

In this section we present results under the Tversky and Kahneman (1992) specification of the value and weighting function (see (1.1) and (1.3)) with $\alpha_\pm = \alpha$ and $\delta_\pm = \delta$. We consider $T = 2$ a two-period model and a starting net wealth at the reference point which is assumed to be zero.

[9] It is also possible to understand the optimal strategy of this type of agent in a game-theoretic manner. The stopping problem can be interpreted as a sequential game with T players, each of them representing a copy of the agent at each time point $t = 0, 1, ..., T-1$. The optimal strategy can then be characterised as a sub-game perfect equilibrium.

The agent who can pre-commit

Consider first Figure 2.1. The left panels show the strategies followed by the agent in the case where he cannot randomise, for three different parameter combinations. An open circle at a node indicates a node at which it is optimal to continue gambling; a closed or solid circle indicates a node at which it is optimal to stop. We find three possible behaviours - the loss-exit strategy of gambling at time zero and continuing after a win but stopping after a loss [Panel (a)]; the gain-exit strategy of gambling at time zero and continuing after a loss but stopping after a win [Panel (c)] and the trivial strategy of always stopping such that the agent does not enter the game in the first place [Panel (e)]. Note that it is never optimal to always continue. The right hand panels show what happens if, for the same parameter sets, we consider an agent who can randomise. We find cases where the agent who can randomise gambles at time zero when the agent who cannot does not gamble. The probability of gambling can be significant in this case, for the parameters of Panels (e) and (f) it is 0.47 as opposed to zero. However, compare Panels (a) and (b) ((c) and (d)), if the deterministic agent follows a loss-exit (gain-exit) strategy then his counterpart with the ability to randomise follows a strategy with similar characteristics.

Now consider the results of Figure 2.2 which describe the strategies in detail as parameters change. Panel (a) shows the parameter combinations (in loss aversion k and α, we fix the probability weighting parameter δ) for which the agent gambles at time zero. The agent without the ability to randomise follows a gain-exit strategy for low values of loss aversion k and convexity/concavity α (region bounded by the dash-dot line) and a loss-exit strategy for low k and large α (region bounded by the solid line) and chooses never to gamble for large values of loss aversion k (the region above the solid and dash-dot lines). Also plotted in Panel (a) by the dashed line is the boundary of the region where the agent with the ability to randomise chooses a non-zero probability of gambling. (Again, the region below the dashed line is the parameter region where he follows a non-trivial strategy). Note that the parameter region where the agent who can randomise enters the gamble at time zero is larger than the corresponding region for the deterministic agent. In this sense, the ability to randomise leads to more gambling.

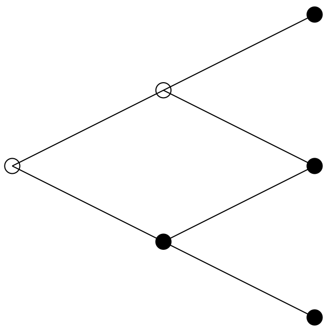
The lower left panel (c) shows the optimal probability that the agent gambles at time zero for the agent who can randomise. The two right panels (b) and (d) show the probability of gambling at time 1, after a win (Panel (b)) or after a loss (Panel (d)). In cases where the agent who cannot randomise follows a gain-exit strategy then the agent who can randomise follows a similar strategy - he gambles at time zero and again after a loss, but after a win he may gamble again, albeit with a small probability. In cases where the agent who cannot randomise follows a loss-exit strategy then the agent who can randomise also exits after a loss, but gambles at time zero with a probability typically less than one, and sometimes as low as 0.4, and also gambles after a win with a probability less than one.

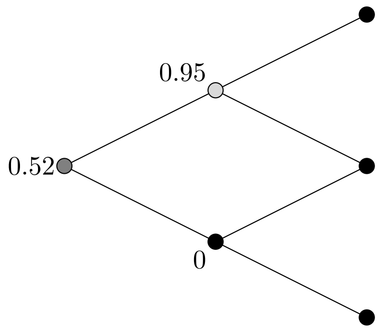
There are several features of PT which drive these results. First, the presence of loss aversion means that symmetric bets, such as stopping at time 1 in all situations, are unattractive. (Further, increasing loss aversion parameter k reduces the value of all strategies making stopping at time zero more likely.) Second, the fact he is risk seeking on losses and risk averse on gains means an agent has an incentive to follow a gain-exit strategy. Since $U(2)/U(1) = U(-2)/U(-1) = 2^{\alpha}$ this incentive is greatest when α is small. However the impact of probability weighting is to provide incentives in the opposite direction - probability weighting gives greater prominence to extreme events, encouraging long tailed distributions on gains and thin tailed distributions on losses, or equivalently, encouraging loss-exit strategies. When α is large and thus the impact of convexity is small, this factor dominates.

What then is the impact of randomisation? In this model randomisation gives the agent further flexibility in the design of his terminal wealth. This makes it more likely that the agent gambles or enters the casino initially, see Panel (a) of Figure 2.2. For the agent who is not allowed to make use of randomisation, the only values of w which are relevant are $w(1/4)$ and $w(1/2)$. In particular, there are no events of very small probability and the shape of w near zero is irrelevant. In contrast, the agent who can randomise can design gambles of arbitrary probability.

The naive agent

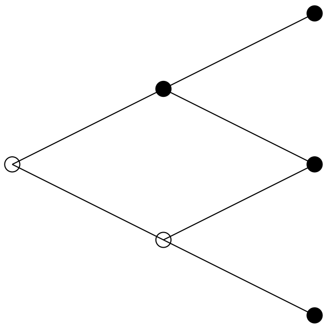
Results for the naive agent are presented in Figures 2.3 and 2.4 in the same format. Recall that we are assuming a version of naïvité whereby the agent can commit his current self to following an action which depends on a contemporaneous random outcome. At time zero, the naive agent selects the identical strategy to the precommiting agent. That is, what he *plans* to do is the same as the agent who can commit. (See the time zero node in all trees in Figures 2.1 and 2.3 and the left hand panels of Figures 2.2 and 2.4.) The interesting feature of the naive agent is the fact that because he is unable to commit, he has the possibility of modifying his strategy at time 1. The naive agent re-evaluates his strategies at time 1, and always chooses to gamble after a loss but stop after a win, ie. if he gambles at time zero, he follows a gain-exit strategy. As Barberis (2012) observes, the striking feature is

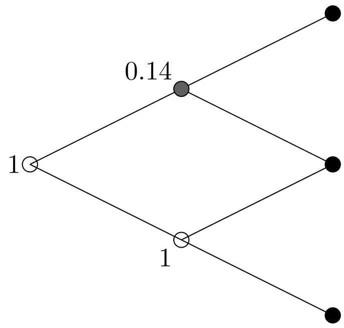




(a) Optimal pure strategy with $\alpha =$ 0.95, $k = 1.45$ and $\delta = 0.5$.

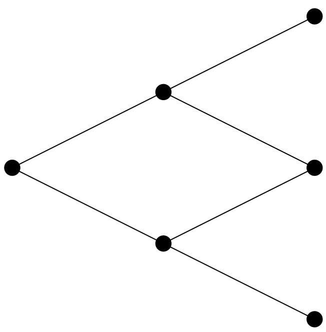
(b) Optimal randomised strategy with $\alpha = 0.95$, $k = 1.45$ and $\delta = 0.5$.

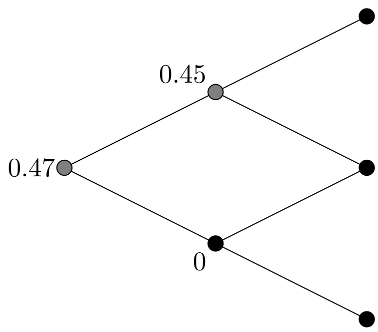




(c) Optimal pure strategy with $\alpha = 0.3$, $k = 1.05$ and $\delta = 0.5$.

(d) Optimal randomised strategy with $\alpha = 0.3$, $k = 1.05$ and $\delta = 0.5$.

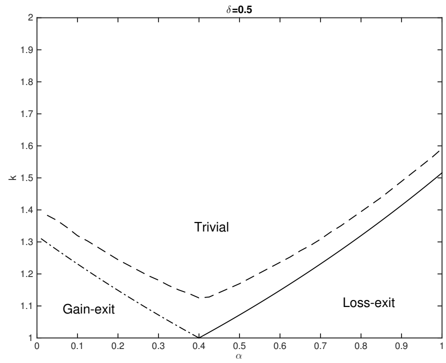




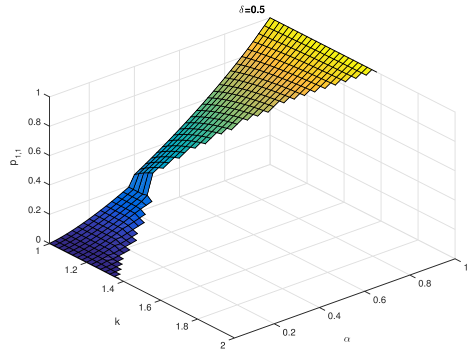
(e) Optimal pure strategy with $\alpha = 0.5$, $k = 1.15$ and $\delta = 0.5$.

(f) Optimal randomised strategy with $\alpha = 0.5$, $k = 1.15$ and $\delta = 0.5$.

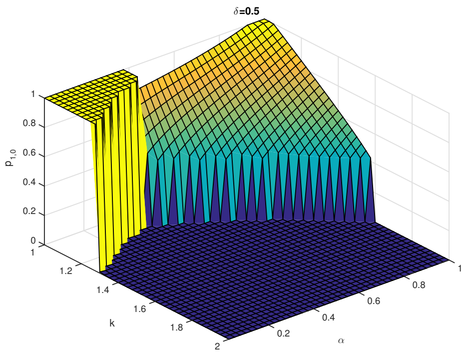
Figure 2.1: *The agent who can precommit.* The left panels display strategies when the agent cannot randomise where an open circle denotes a node at which it is optimal to continue and closed or solid circles are nodes where it is optimal to stop. The right panels display the corresponding strategy for the agent who can randomise where the number at each node gives the *probability of continuing* at that node. The Tversky and Kahneman (1992) PT parameters are given in each panel.



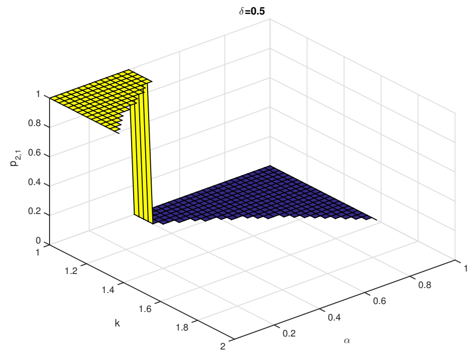
(a) Parameter combinations for which the agent gambles at time zero and form of his strategy.



(b) Probability of continuing at the gain node for an agent with randomised strategies.



(c) Probability of entering the game at time zero for an agent using randomised strategies.



(d) Probability of continuing at the loss node for an agent with randomised strategies.

Figure 2.2: *The agent who can precommit.* Panel (a) displays the parameter combinations for which the agent gambles at time zero. The solid and dash-dot lines distinguish the three types of strategy {do not enter/trivial; loss-exit, gain-exit} for the agent without randomisation. The dashed line indicates the region below which the agent who can randomise enters the game at time zero with some positive probability. Panels (c) displays the probability of entering the game at time zero for the agent who can randomise. Panels (b) (and (d)) display the probability of continuing at the gain (loss) node at time 1 for the agent who can randomise. Regions with no value in (b) and (d) correspond to parameters where the agent did not enter the game at time zero. All panels use $T = 2$ and $\delta = 0.5$.

that when the probability weighting is the dominant effect, the naive agent without access to randomisation plans at time zero to follow a loss-exit strategy, but actually follows a gain-exit strategy. For the agent with randomisation, the results are broadly similar. The impact of randomisation is to modify the probability of entering the casino initially, but at time 1, this agent still follows a gain-exit strategy.

The sophisticated agent

The results for the sophisticated agent are obtained by backward induction, see Figure 2.5. In comparison with the results for the agent with the ability to commit, we find in the region where the latter follows a gain-exit strategy the two agents behave similarly, but in the region where the latter follows a loss-exit strategy, the sophisticated agent without the ability to commit follows the trivial strategy. The agent knows at time zero that he plans to follow a strategy of gain-exit at time 1 (if he does not follow the trivial strategy). Hence, in regions where the agent with precommitment would choose to follow a loss-exit strategy (where loss aversion and convexity/concavity are small and the probability weighting dominates) the sophisticated agent knows that his future self will not follow this strategy and prefers not to gamble at time zero.

2.4 The continuous model

2.4.1 The setup of Ebert and Strack (2015)

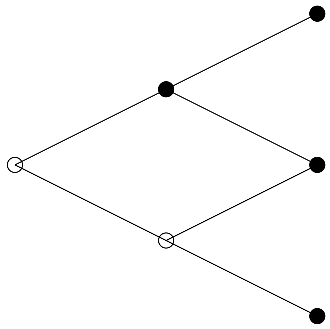
In the Ebert and Strack (2015) model, returns from gambling are modelled as a Brownian motion B^x where the superscript x denotes the starting wealth level. The agent's problem is thus to find

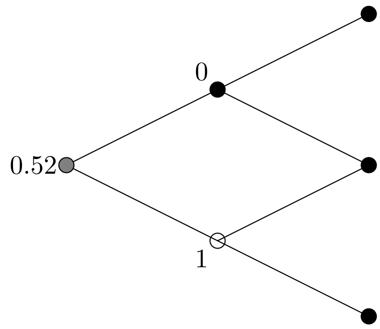
$$\tilde{V}(x) = \sup_{\tau \in \Lambda} \mathcal{E}(B^x_\tau) \tag{2.1}$$

where Λ is the set of uniformly integrable stopping times.[10]

The uniform integrability assumption deserves some discussion. Since Brownian motion hits any level in finite time with probability one, over the infinite horizon an agent without a lower bound on wealth can achieve a prospect $(y, 1)$ (a unit mass at y) for any y, especially any $y > x$ where x is the starting value of the Brownian motion. We want to exclude this type of doubling strategy. There are at least four approaches or rationales

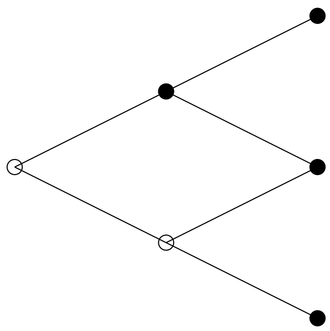
[10]Recall that the set of uniformly integrable stopping times is the set of finite stopping times ρ such that $(B^x_{t\wedge\rho})_{t\geqslant 0}$ is a uniformly integrable family of random variables.

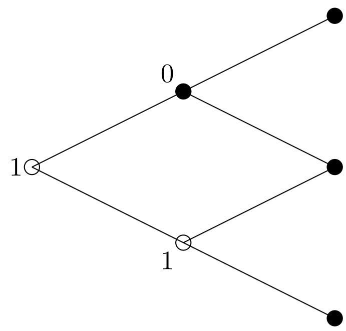




(a) Optimal pure strategy with $\alpha = 0.95$, $k = 1.45$ and $\delta = 0.5$.

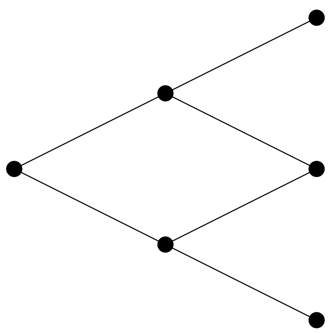
(b) Optimal randomised strategy with $\alpha = 0.95$, $k = 1.45$ and $\delta = 0.5$.

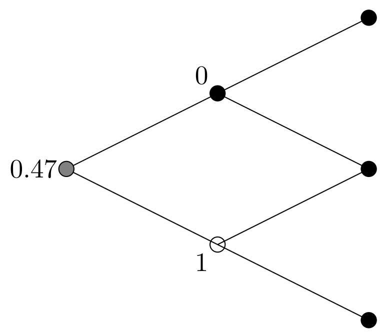




(c) Optimal pure strategy with $\alpha = 0.3$, $k = 1.05$ and $\delta = 0.5$.

(d) Optimal randomised strategy with $\alpha = 0.3$, $k = 1.05$ and $\delta = 0.5$.

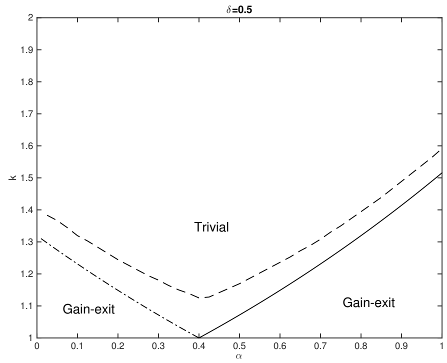




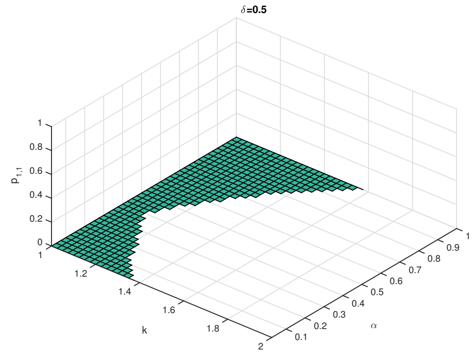
(e) Optimal pure strategy with $\alpha = 0.5$, $k = 1.15$ and $\delta = 0.5$.

(f) Optimal randomised strategy with $\alpha = 0.5$, $k = 1.15$ and $\delta = 0.5$.

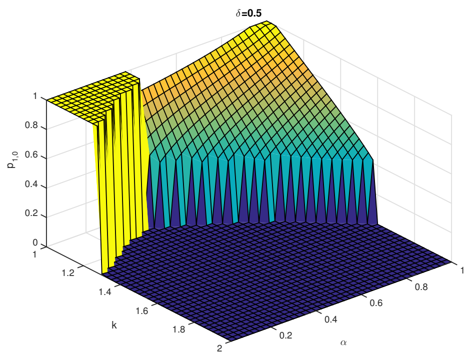
Figure 2.3: *The naive agent.* The left panels display strategies when the agent cannot randomise where an open circle denotes a node at which it is optimal to continue and closed or solid circles are nodes where it is optimal to stop. The right panels display the corresponding strategy for the agent who can randomise where the number at each node gives the *probability of continuing* at that node. The Tversky and Kahneman (1992) PT parameters are given in each panel.



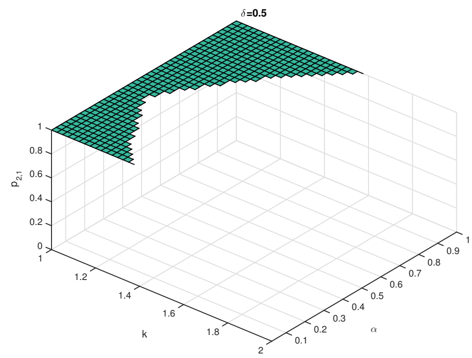
(a) Parameter combinations for which the agent gambles at time zero and form of his strategy.



(b) Probability of continuing at the gain node for an agent with randomised strategies.



(c) Probability of entering the game at time zero for an agent using randomised strategies.



(d) Probability of continuing at the loss node for an agent with randomised strategies.

Figure 2.4: *The naive agent.* Panel (a) displays the parameter combinations for which the agent gambles at time zero. The solid and dash-dot lines distinguish the two types of strategy {do not enter/trivial; gain-exit} for the agent without randomisation. The dashed line indicates the region below which the agent who can randomise enters the game at time zero with some positive probability. Panels (c) displays the probability of entering the game at time zero for the agent who can randomise. Panels (b) (and (d)) display the probability of continuing at the gain (loss) node at time 1 for the agent who can randomise. Regions with no value in (b) and (d) correspond to parameters where the agent did not enter the game at time zero. All panels use $T = 2$ and $\delta = 0.5$.

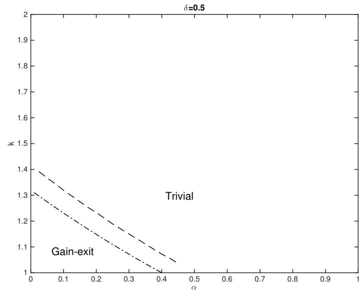


Figure 2.5: *The sophisticated agent.* The figure displays the parameter combinations for which the sophisticated agent gambles at time zero. The dash-dot line distinguishes the two types of strategy {do not enter/trivial; gain-exit} for the agent without randomisation. The dashed line indicates the region below which the agent who can randomise enters the game at time zero with some positive probability. In the region where the agent enters, the probabilities of continuing at the gain (loss) nodes are zero (one) respectively. We use $T = 2$ and $\delta = 0.5$.

for doing so. First, we can suppose that the Brownian motion has a negative drift (so that Brownian motion B^x hits levels $y > x$ with probabilities strictly less than one), at the expense of an extra level of complication in the analysis. Second, we can insist that if the Brownian motion hits some value $-\kappa$ then all gambling must cease, and the agent receives the stopped value $-\kappa$. It follows that B^x is a supermartingale and all feasible prospects (ie. the class of random variables in form of B^x_τ) have expected values less than or equal to x. Then by first order stochastic dominance it is sufficient to consider only prospects with mean x. Having calculated the optimal prospect we can check that its range is bounded below, and hence that the condition of enforced stopping at $-\kappa$ is never applied, for large enough κ. Third, we could insist that stopping times are bounded by some large $\overline{T}$ and then let $\overline{T}$ increase to infinity. Finally, as we do, we could exclude doubling strategies by positing a mathematical restriction on τ, in particular by insisting that stopping times are uniformly integrable, whence the mean of the stopped process is x.

(2.1) is the type of the probability-weighted optimal stopping problem that we analyse in Chapter 1. The key step towards constructing a solution is based on the idea that the search over stopping times can be replaced by a search over random variables (refer to the equivalence of (1.7) and (1.8) in Section 1.2.2 of Chapter 1). Define

$$V(x) = \sup_{X:\mathbb{E}[X]=x} \mathcal{E}(X).$$

Then we have $V(x) = \tilde{V}(x)$. Our goal here is to find $V(x) = \tilde{V}(x)$ and the set of optimisers for V (prospects or random variables) and $\tilde{V}$ (stopping times).

Ebert and Strack (2015) show that for an extremely wide class of value and probability weighting functions, for any initial wealth x there is a first exit stopping rule (a two-point prospect) which is preferred to stopping immediately (a prospect of a sure-wealth x). From this they reach the result that in continuous time models, naive agents always postpone their stopping decisions: a naive agent with PT preferences will gamble *"until the bitter end"*. Taken at face value, this never stopping result leads to unrealistic predictions and casts doubt on the usefulness of naïvité plus PT preferences in a dynamic context.

Assume the reference level is zero, and that the initial value of the Brownian motion is at the reference level. We first restate a result of Ebert and Strack (2015).

Proposition 2.1 (Ebert and Strack (2015))**.** *Suppose that (in addition to the standard assumptions listed in Section 2.2.1) the value function U satisfies $0 < U'(0-) = KU'(0+) < \infty$ for some $K > 1$.*[11,12]

Suppose for the probability weighting functions $w_\pm$ there exists a $\hat{p} \in (0, 1/2)$ such that

$$w_+(\hat{p}) > \frac{K\hat{p}}{1+(K-1)\hat{p}}, \qquad w_-(1-\hat{p}) < \frac{1-\hat{p}}{1+(K-1)\hat{p}}. \tag{2.2}$$

Then there exists $\hat{\epsilon} > 0$ such that the agent with zero initial wealth prefers gambling until his wealth reaches $\hat{\epsilon}(1-\hat{p})$ or $-\hat{\epsilon}\hat{p}$ to stopping immediately.

Proof. Let ϵ be a positive constant which later we will treat as a parameter. Let $b_\epsilon = \epsilon(1-\hat{p})$ and $a_\epsilon = -\epsilon\hat{p}$. Suppose the agent has zero initial wealth. One strategy open to the agent is to gamble until the first time his wealth reaches b_ϵ or a_ϵ and then to stop. By construction, the probability that the process hits b_ϵ before a_ϵ is $\hat{p}$. Then the value $H_{\hat{p}}(\epsilon)$ of this strategy is:

$$H_{\hat{p}}(\epsilon) = w_+(\hat{p})U(b_\epsilon) + w_-(1-\hat{p})U(a_\epsilon) = w_+(\hat{p})U(\epsilon(1-\hat{p})) + w_-(1-\hat{p})U(-\epsilon\hat{p}).$$

[11] Under this assumption the value function exhibits finite marginal loss aversion in the sense that $U'(0+) < U'(0-) < \infty$. Köbberling and Wakker (2005) introduce this notion of loss aversion and show that it has the advantage of being scale independent. We have added the adjective marginal to distinguish the concept from simple loss aversion $U(y) + U(-y) < 0$.

[12] Note that the Kahneman and Tversky (1979) value function (1.1) has infinite slope at the origin and thus does not satisfy this assumption. We postpone the discussion on this type of value function to Section 2.6.

		Tversky-Kahneman	Goldstein-Einhorn	Prelec
Condition (2.2)	$K = 2.25$	(0, 0.072)	(0, 0.035)	(0, 0.062)
	$K = 1.5$	(0, 0.176)	(0, 0.104)	(0, 0.139)
Condition (2.3)		(0, 0.5)	(0, 0.5)	(0.0018, 0.5)

Table 2.1: Ranges of probabilities for which the Ebert-Strack condition (2.2) and condition (2.3) are satisfied for various weighting function specifications and input parameters. The first two rows correspond to those in Table W.1 of Ebert and Strack (2015) and give probabilities for which (2.2) is satisfied. The third row gives the range of probabilities for which condition (2.3) is satisfied. The first (second; third) column corresponds to the Tversky and Kahneman (1992) weighting function with $\delta_\pm = 0.65$ (Goldstein and Einhorn (1987) with $d_\pm = 0.69, \gamma_\pm = 0.77$; Prelec (1998) with $b_\pm = 1.05, a_\pm = 0.65$).

Note that $H_{\hat{p}}(0) = 0$. Now, writing $H'_{\hat{p}}$ for the derivative with respect to ϵ,

$$\begin{aligned} H'_{\hat{p}}(0+) &= (1-\hat{p})w_+(\hat{p})U'(0+) - \hat{p}w_-(1-\hat{p})U'(0-) \\ &> \hat{p}(1-\hat{p})\left\{\frac{K}{1+(K-1)\hat{p}} - \frac{K}{1+(K-1)\hat{p}}\right\}U'(0+) = 0. \end{aligned}$$

Hence there exists $\hat{\epsilon} > 0$ for which $H_{\hat{p}}(\hat{\epsilon}) > 0$ and for this $(\hat{p}, \hat{\epsilon})$ the agent prefers to continue (run until wealth first hits $b_{\hat{\epsilon}}$ or $a_{\hat{\epsilon}}$) over stopping immediately.

□

From this result (together with an extension to any arbitrary initial wealth level x), Ebert and Strack (2015) infer that a naive agent who can only use pure stopping rules always postpones stopping decisions and hence, never stops.

The assumptions on the weighting functions given in (2.2) are satisfied by the commonly used inverse-S shaped weighting functions of Tversky and Kahneman (1992), Prelec (1998), Goldstein and Einhorn (1987) (see (1.4) for the forms of the Prelec and Goldstein-Einhorn weighting functions) and the neo-additive weighting function (Wakker (2010), p208) for many parameter values, provided $\hat{p}$ is small. The first two rows of Table 2.1 correspond to those in Table W.1 of Ebert and Strack (2015) and give probabilites for which (2.2) is satisfied. For example, for the Tversky and Kahneman (1992) weighting function, and a loss aversion parameter of $K = 2.25$, (2.2) is satisfied provided the gain probability $\hat{p}$ is chosen to be less than 0.072.

Ebert and Strack (2015) show that a sufficient condition for there to exist a $\hat{p}$ such that (2.2) holds is both $w'_+(0+) > K$ and $w'_-(1-) > K$. This property says that extremely unlikely gains are overweighted and extremely likely losses are underweighted, both by more than the loss aversion parameter. That is, *probability weighting is stronger than loss aversion.*

2.4.2 Randomisation in the continuous model

Now we move on to mixed strategies and our first main result in the continuous time setup. As in the discrete model, the fact that a PT agent prefers one strategy over another does not mean that he necessarily prefers the first strategy over any mixture of the two. We now show that in the setup considered in the main body of Ebert and Strack (2015), the agent who can randomise his strategy prefers such a mixture. Thus if the naive agent can mix over strategies then he prefers sometimes stopping over never stopping. This prediction is more realistic and is closer to observed behaviour.

Proposition 2.2. *Suppose that in addition to the assumptions of Proposition 2.1 we have*

$$w'_-((1-\hat{p})) > w'_+(\hat{p}). \tag{2.3}$$

Then there exists $\hat{\theta} \in (0,1)$ such that the agent with zero initial wealth prefers a randomised strategy of stopping immediately with probability $(1-\hat{\theta})$ and otherwise gambling until his wealth reaches $\hat{\epsilon}(1-\hat{p})$ or $-\hat{\epsilon}\hat{p}$ to the pure strategy of gambling until his wealth reaches $\hat{\epsilon}(1-\hat{p})$ or $-\hat{\epsilon}\hat{p}$.

Proof. We use the same notation as in the proof of Proposition 2.1. Suppose now the agent stops gambling immediately with probability $1-\theta$ and otherwise gambles until his wealth reaches $b_{\hat{\epsilon}}$ or $a_{\hat{\epsilon}}$. Fixing $\hat{p}, \hat{\epsilon}$, considering θ as a variable and writing $H_{\hat{p},\hat{\epsilon}}(\theta)$ as the value of the strategy,

$$H_{\hat{p},\hat{\epsilon}}(\theta) = w_+(\theta\hat{p})U(b_{\hat{\epsilon}}) + w_-(\theta(1-\hat{p}))U(a_{\hat{\epsilon}}) = w_+(\theta\hat{p})U(\hat{\epsilon}(1-\hat{p})) + w_-(\theta(1-\hat{p}))U(-\hat{\epsilon}\hat{p}).$$

Note that $H_{\hat{p},\hat{\epsilon}}(0) = 0$ and $H_{\hat{p},\hat{\epsilon}}(1) > 0$ by design.

Consider the derivative of H with respect to θ. Then

$$\begin{aligned}
\frac{\partial}{\partial\theta} H_{\hat{p},\hat{\epsilon}}(\theta) &= \hat{p}w'_+(\theta\hat{p})U(\hat{\epsilon}(1-\hat{p})) + (1-\hat{p})w'_-(\theta(1-\hat{p}))U(-\hat{\epsilon}\hat{p}) \\
&= \hat{p}(1-\hat{p})\left[w'_+(\theta\hat{p})\frac{U(\hat{\epsilon}(1-\hat{p}))}{1-\hat{p}} + w'_-(\theta(1-\hat{p}))\frac{U(-\hat{\epsilon}\hat{p})}{\hat{p}}\right].
\end{aligned}$$

Since U is concave on $y > 0$, $U(y) + U(-y) < 0$ and $\hat{p} < 1/2$, we have

$$\frac{U(-\hat{\epsilon}\hat{p})}{\hat{p}} < -\frac{U(\hat{\epsilon}\hat{p})}{\hat{p}} < -\frac{U(\hat{\epsilon}(1-\hat{p}))}{1-\hat{p}}.$$

Recall our hypothesis that $w'_-((1-\hat{p})) > w'_+(\hat{p})$. Then

$$\frac{\partial}{\partial\theta} H_{\hat{p},\hat{\epsilon}}(\theta)\bigg|_{\theta=1} < \hat{p}\left\{w'_+(\hat{p}) - w'_-(1-\hat{p})\right\} U(\hat{\epsilon}(1-\hat{p})) < 0.$$

In particular, $H_{\hat{p},\epsilon}(\theta)$ is maximised at some interior point and the agent who can randomise prefers a mixed strategy of sometimes stopping and sometimes waiting until wealth first leaves the interval $(a_{\hat{\epsilon}}, b_{\hat{\epsilon}})$ to a pure strategy of waiting until wealth first leaves the same interval.

□

Intuitively, condition (2.3) says we require the probability weighting function on losses for nearly certain events to be stronger than probability weighting on gains for rare events. Since typical weighting functions satisfy $w_\pm(1/2) < 1/2$, $w_\pm$ is steeper over $(1/2, 1)$ than over $(0, 1/2)$ and hence (2.3) may be expected to hold for a wide family of popular weighting functions. Since $\int_0^{1/2}[w'_-((1-p)) - w'_+(p)]dp = 1 - w_-(1/2) - w_+(1/2) > 0$ there must exist a range of $\hat{p}$ for which $w'_-((1-\hat{p})) > w'_+(\hat{p})$.

We now discuss specific functional forms in the literature. By considering $\lim_{p\downarrow 0} \frac{w'_-(1-p)}{w'_+(p)}$ we see that for the Tversky and Kahneman (1992) weighting functions, (2.3) is satisfied for all sufficiently small $\hat{p}$ provided $\delta_+ \geq \delta_-$. For the weighting function of Goldstein and Einhorn (1987), (2.3) is satisfied for all sufficiently small $\hat{p}$, provided $d_+ > d_-$ or $d_+ = d_-$ and $\gamma_+ \geq \gamma_-$. In the third row of Table 2.1 we report the range of probabilities for which (2.3) is satisfied. We take the same parameters as used by Ebert and Strack (2015) in their Table W.1. We see that for their parameter choices, condition (2.3) holds for all probabilities $\hat{p} < 1/2$ for the Tversky and Kahneman (1992) and Goldstein and Einhorn (1987) weighting functions. In particular, the condition (2.3) is very mild and is easily satisfied for these parameters. There is also a column in Table 2.1 for the Prelec (1998) weighting function. Although (2.3) is not always satisfied by the Prelec (1998) weighting function, it is satisfied for almost the whole range of probabilities for the parameters used by Ebert and Strack (2015), excluding only very small values of $\hat{p}$ less than 0.0018.

All the examples in Table 2.1 assume $w_+ = w_-$. An example with $w_+ \neq w_-$ is the Tversky and Kahneman (1992) weighting function with parameters of $\delta_+ = 0.61, \delta_- = 0.69$ and loss aversion $K = 2.25$. Then Ebert and Strack's condition (2.2) holds for $\hat{p} < 0.0839$

and condition (2.3) holds over the range $(0.0023, 0.5)$. In particular, both (2.2) and (2.3) hold for $0.0023 < \hat{p} < 0.0839$.

Proposition 2.2 shows that the first exit strategy of Ebert and Strack (2015) can be improved upon by mixing with sometimes stopping immediately, and hence their argument is not sufficient to conclude that a naive agent never stops if he has access to randomised strategies. Instead, we need to identify the optimal prospect: if this prospect has no mass at the starting wealth level, then a naive agent with the ability to randomise would never stop. Our goal in the next section is to show that in general this is not the case, by giving examples in which the optimal prospects have masses at zero.

2.5 Two stylised examples

In this section, we give simple, tractable examples which demonstrate that the optimal prospect can have a mass at the starting wealth level. As before, the wealth process prior to stopping is a Brownian motion starting at the reference level which is assumed to be zero. We consider two different specifications of the value function U and probability weighting functions $w_{\pm}$. Using (1.8) and a change of variable, our problem in both cases is to find (the optimiser for)

$$V = \sup\left\{\int_0^\infty w_+(\bar{F}_X(x))U'(x)dx - \int_{-\infty}^0 w_-(F_X(x))U'(x)dx\right\} \tag{2.4}$$

where the supremum is taken over random variables X with mean 0. F_X is the CDF of X and $\bar{F}_X(x) := 1 - F_X(x)$.

2.5.1 Specifications

Specification I

In our first specification, the value function takes the piecewise linear form

$$U(x) = \begin{cases} x \wedge 1, & x \geq 0, \\ Kx, & x < 0, \end{cases}$$

where $K > 1$. Then U exhibits loss aversion (both simple and finite marginal) and is concave over gains[13] and convex on losses.

[13]Some form of concavity is required on gains, else the optimisation problem for the PT agent with probability weighting is ill-posed. The cap on the value function for wealth above one is a particularly simple form, and may represent the not unrealistic assumption that the agent will be asked to leave the casino once his cumulative winnings reach a critical value.

The probability weighting functions $w_\pm$ are general inverse-S shaped satisfying the standard assumptions in Section 2.2.1. Define $\eta = \min_{p\in[0,1]} \frac{w_-(p)}{p}$ and $q_- = \operatorname{argmin}_{p\in[0,1]} \frac{w_-(p)}{p}$. Also set

$$q_+ = \operatorname*{argmax}_{p\in[0,1]} \left\{ w_+(p) - K\eta p \right\}. \tag{2.5}$$

We further assume that

$$q_+ + q_- = \operatorname*{argmax}_{p\in[0,1]} \left\{ w_+(p) - K \min_{u\in[0,1]} \left(\frac{w_-(u)}{u} \right) p \right\} + \operatorname*{argmin}_{p\in[0,1]} \left\{ \frac{w_-(p)}{p} \right\} < 1. \tag{2.6}$$

If $w_+ = w_- = w$ and the probability weighting function is that of Tversky and Kahneman (1992) then (2.6) is satisfied provided there is a reasonable level of loss aversion. For example if $\delta = 0.65$, (2.6) is satisfied by K greater than about 1.2. See Appendix 2.A for a pair of simple sufficient conditions for (2.6). If $K = 1$ such that there is no loss aversion, then (2.6) cannot hold.

Specification II

In the second case, the value function is logarithmic over gains and linear on losses taking the form of

$$U(x) = \begin{cases} \ln(1+x), & x \geq 0, \\ Kx, & x < 0. \end{cases} \tag{2.7}$$

As before, we impose simple as well as finite marginal loss aversion by setting $K > 1$.

The weighting function on losses w_- is general inverse-S shaped. We again define $\eta = \min_{p\in[0,1]} \frac{w_-(p)}{p}$ and $q_- = \operatorname{argmin}_{p\in[0,1]} \frac{w_-(p)}{p}$. Meanwhile, the weighting function on gains is piecewise linear given by

$$w_+(p) = \begin{cases} \alpha p, & 0 \leq p \leq q_+, \\ \alpha q_+ + \beta(p - q_+), & q_+ < p < \frac{\alpha-1}{\alpha-\beta} + q_+, \\ 1 - \alpha(1-p), & \frac{\alpha-1}{\alpha-\beta} + q_+ \leq p \leq 1, \end{cases} \tag{2.8}$$

where $\alpha > 1$, $\beta \in (0,1)$ and q_+ are constants with[14] $0 < q_+ < \frac{1-\beta}{2(\alpha-\beta)} < \frac{1}{2}$. Piecewise linear weighting functions of this form are proposed by Webb (2015) as a simple generalisation of the probability weighting function used in the NEO-expected utility of Chateauneuf et al. (2007). Relative to the NEO-expected utility probability weighting functions, a piecewise linear function has the advantage of being continuous. Figure 2.6 gives a sketch of w_+ and w_- under this specification.

[14] A condition $q_+ \leq \frac{1-\beta}{\alpha-\beta}$ is required to ensure that $\frac{\alpha-1}{\alpha-\beta} + q_+ \leq 1$. The additional factor of $\frac{1}{2}$ ensures that $w_+(\frac{1}{2}) < \frac{1}{2}$.

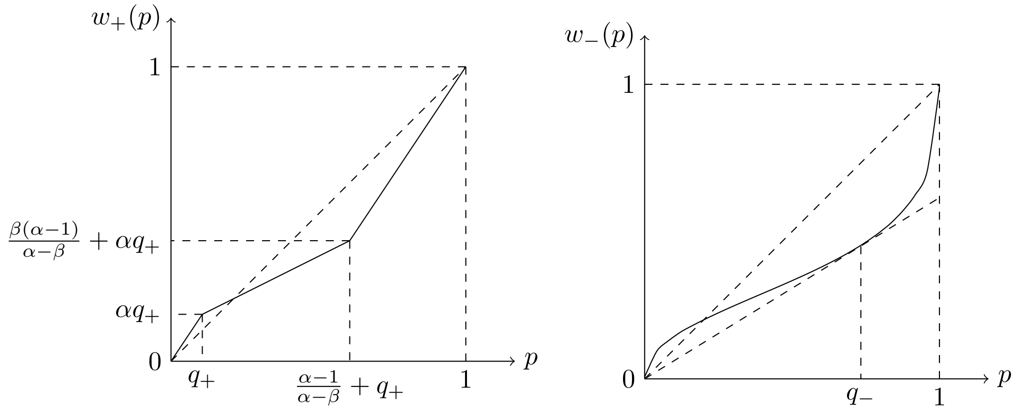


Figure 2.6: Sketch of w_+ and w_- used in Specification II of Section 2.5. w_- is a general inverse-S shaped weighting function whilst w_+ is a piecewise linear function defined in (2.8).

Finally, we assume $q_+ + q_- < 1$ (as in (2.6) under Specification I) and

$$\frac{1-\alpha q_+}{1-q_+} < K\eta < \alpha. \tag{2.9}$$

Note that under assumption (2.9) and using the piecewise linear structure of w_+, we have

$$\max_{p\in[0,1]} \{w_+(p) - K\eta p\} = \max\left(0, 1-K\eta, q_+(\alpha - K\eta)\right) = q_+(\alpha - K\eta)$$

and the optimiser is q_+. Hence $q_+ = \operatorname{argmax}_{p\in[0,1]} \{w_+(p) - K\eta p\}$. Even though q_+ should be interpreted as a given model parameter under this specification, it is consistent with definition (2.5) under Specification I.

2.5.2 The optimal prospects in the stylised examples

Proposition 2.3. *If the initial wealth level is $x = 0$, then the optimal prospects $\mathcal{P}^* = \mathcal{P}_0^*$ under Specification I and II share the same form of a three-point distribution as*

$$\mathcal{P}^* = \left(-\frac{\nu}{q_-}, q_-; 0, 1-q_- - q_+; \frac{\nu}{q_+}, q_+\right) \tag{2.10}$$

where

$$q_- = \operatorname*{argmin}_{p\in[0,1]} \frac{w_-(p)}{p}, \qquad q_+ = \operatorname*{argmax}_{p\in[0,1]} \left\{ w_+(p) - K \min_{u\in[0,1]} \left(\frac{w_-(u)}{u}\right) p \right\}.$$

Moreover, $\nu = q_+$ under Specification I and $\nu = q_+\left(\frac{\alpha q_-}{K w_-(q_-)} - 1\right)$ under Specification II.

The proof is given in Appendix 2.B. It follows that for an initial wealth starting at the reference level of zero the optimal prospect includes a point mass at the reference level.

If, in addition to the restrictions on parameters of $w_\pm$ in Section 2.5.1, we add an assumption that there exists some $\overline{p}$ such that (2.2) and (2.3) hold for all $\hat{p} < \overline{p}$.[15] Then we have examples which satisfy the conditions of Proposition 2.1, Proposition 2.2 and Proposition 2.3. For these examples, a naive agent with no access to randomisation and with zero initial wealth will never stop (Proposition 2.1, Ebert and Strack (2015)). Such an agent can benefit from randomisation to improve the first exit strategy of Ebert and Strack (2015) (Proposition 2.2). His optimal prospect includes an atom at zero (Proposition 2.3) and we discuss the possible optimal behaviours of the naive agent in Section 2.5.4.

2.5.3 The optimal stopping rules in the stylised examples

The conclusion from the previous section is that there are circumstances in which the optimal prospect includes a mass at the origin. Typically (as in the example above) there is uniqueness at the level of optimal prospects. But there are many stopping rules an agent might use to attain a given prospect. In the probability literature, these are known as solutions of the Skorokhod embedding problems.[16]

For now, assume that the agent is a precommitting one and he is looking for a stopping strategy which can attain the target prospect $\mathcal{P}_0^*$. One way to do so is to use a stopping rule in which the stopping time is the first time that the Brownian motion falls below a well-chosen function of its running maximum. The Azéma and Yor (1979) solution of the Skorokhod embedding problem takes this form.

A second way to achieve the prospect $\mathcal{P}_0^*$ is to wait until Brownian motion hits $\hat{a} = -\nu \frac{q_+ + q_-}{q_-}$ or $\hat{b} = \nu \frac{q_+ + q_-}{q_+}$, and then to stop the first time thereafter that the Brownian motion is at $-\frac{\nu}{q_-}$, 0 or $\frac{\nu}{q_+}$. For an agent with access to a randomisation device (in the form of an independent random variable), other stopping rules can be used to attain optimality. The simplest is to stop immediately with probability $1 - q_+ - q_-$ and to otherwise stop the first time the Brownian motion reaches $-\frac{\nu}{q_-}$ or $\frac{\nu}{q_+}$.

A further way to achieve the prospect $\mathcal{P}_0^*$ is to stop the first time the Brownian motion hits $-\frac{\nu}{q_-}$ or $\frac{\nu}{q_+}$, and also to stop at zero at a constant rate $\Phi := \frac{1-q_+-q_-}{2\nu}$. Then,

[15] In both Specification I and II, it is possible to construct examples of $w_\pm$ which satisfy all these conditions simultaneously. See Appendix 2.C.

[16] The only cases where there is a unique embedding (within the class of uniformly integrable embeddings) are those where the optimal prospect is either a point mass at the reference level (when stopping immediately is the only strategy) or a pair of point masses at locations $\breve{a} < 0 < \breve{b}$ (when the only strategy is to stop on first exit from the interval $(\breve{a}, \breve{b})$). In general, for dynamic optimal stopping problems under probability weighting we must expect non-uniqueness of the optimal strategy.

in the absence of stopping at $-\frac{\nu}{q_-}$ or $\frac{\nu}{q_+}$, the probability that the process has not been stopped by time t is $e^{-\Phi L_t^0}$ where $(L_u^0)_{u\geq 0}$ is the local time of Brownian motion at zero. More specifically, this is the stopping rule introduced in Lemma 1.10 of Chapter 1 with the choice of λ such that $\lambda(\{0\}) = \Phi$, $\lambda(\{-\frac{\nu}{q_-}\}) = \lambda(\{\frac{\nu}{q_+}\}) = \infty$ and zero elsewhere. This strategy is time-homogeneous and Markovian in the sense that the decision to stop only depends on the current value of the wealth process. In fact it is the unique stopping rule with these properties.[17]

2.5.4 Naive agents in the stylised examples

Proposition 2.3 describes the optimal prospect for the PT agent, and the discussion in Section 2.5.3 describes some optimal strategies for an agent with zero initial wealth who can commit to a strategy or stopping rule. What then is the strategy followed by the naive agent? The strategy of the naive agent is the instantaneous, time zero, element of the planned strategy of the agent who can commit to a stopping rule.

When the Brownian motion is at the origin, there is a family of optimal strategies for the agent who can precommit, and no unique prediction for the behaviour of the naive agent. Provided the optimal prospect includes an atom at zero (Proposition 2.3), any time-zero behaviour at the reference level associated with the possible strategies in Section 2.5.3 (including stopping immediately, always continuing, or stopping at a rate) is consistent with the behaviour of a naive agent.[18]

What might be considered a characteristic of a reasonable strategy for a naive agent? One characteristic would be homogeneity in time, so that the same rule is used to stop or otherwise continue each time the process returns to a given level. A second characteristic would be that the decision to stop, now and in the future, depends on the prevailing level of

[17] In this example where the underlying wealth process is a Brownian motion, we say that a stopping time τ is time-homogeneous and Markovian if there exists a time-independent measure λ such that $\mathbb{P}(\tau > t) = \mathbb{P}(\int_{\mathbb{R}} L_t^u \lambda(du) < T)$ where $(L_t^u)_{t\geqslant 0,u}$ is the local time process of the Brownian motion and T is an independent exponential random variable of rate one. Suppose λ has a density such that $\lambda(dx) = \zeta(x)dx$. Using occupation times formula, we could deduce the hazard rate of this stopping strategy as

$$\lim_{\delta t\downarrow 0} \frac{\mathbb{P}(\tau \in (t, t+\delta t) | \tau > t, (B_s)_{s\leqslant t})}{\delta t} = \zeta(B_t).$$

In particular, the agent is stopping the process at a rate per unit time which only depends on the current level of the Brownian motion but not its historical path nor time. If an agent wants to use this type of time-homogeneous, Markovian stopping rule to attain a given target prospect, the required choice of λ is uniquely given by (1.17) in Chapter 1.

[18] If only pure strategies are deemed admissible, then the Ebert and Strack (2015) result implies that instantaneous stopping is never optimal.

the wealth process and not on any other factors.[19] This suggests that a naive agent should adopt a Markovian strategy. Within the class of time-homogeneous, Markovian stopping rules there is a unique strategy which attains a given prospect. If we insist that the naive agent chooses strategies from this class then there is a unique prediction for his behaviour. Except in trivial cases, the use of a Markovian strategy requires randomisation.

Earlier in Section 2.5.3 we gave four stopping rules an agent with the ability to precommit might use to attain the optimal prospect. The first strategy based on the Azéma-Yor stopping time is Markovian in the pair $(B_t, S_t = \sup_{s \leq t} B_s)_{t \geq 0}$ but it is not Markovian in the wealth process B alone. The second strategy is not time-homogeneous. Although neither Markovian nor time-homogeneous, both these strategies are pure, and never involve stopping immediately. A naive agent basing his strategy on these rules never stops, as predicted by Ebert and Strack (2015).

The third proposed strategy for the agent with a commitment device is to stop immediately with probability $1 - q_+ - q_-$, and otherwise to stop the first time the wealth process leaves the interval $(-\frac{\nu}{q_-}, \frac{\nu}{q_+})$. This requires a randomisation device at $t = 0$. The strategy is time-inhomogeneous since the agent plans to use a different strategy on any future returns to zero. A naive agent following this strategy has a positive probability of stopping *every* time the Brownian motion is at the reference level. Since Brownian motion started at zero returns to zero infinitely often in any positive time interval, the naive agent using this strategy stops the first time the Brownian motion hits zero with probability one. For wealth processes started at the reference level, stopping is immediate with probability one, the exact converse to the Ebert and Strack (2015) result of always waiting to the bitter end.

The fourth strategy requires randomisation as well, but is Markovian and time-homogeneous. The naive agent following this stopping rule stops at the origin at rate Φ. The decision about whether to stop or continue depends only on the current value of wealth, and does not depend on any other factors (including time). The discussion of this section can be summarised by the following result.

Theorem 2.4. *Suppose the naive agent uses a Markovian time-homogeneous strategy. Then, for the examples in Section 2.5.1, the agent may stop at the reference level.*

[19] A naive agent will reconsider any planned decisions. At time zero the agent does not plan to base his strategy on any element of the wealth history except the current value: it is not natural for him to plan to use a rule in which stopping at a future time depends on the behaviour of the wealth process between now and that future time.

2.6 Extension to a more general model

Our stylised examples in Section 2.5 have been designed to facilitate calculations. Nonetheless, they demonstrate that it is not the case that naive agents never stop: instead if agents can randomise their strategy then they may stop voluntarily. For our analysis to be as wide-ranging as Ebert and Strack (2015), we need to allow for more general value functions and more general wealth processes. In particular, we want to consider value functions with infinite marginal utility at the origin which are not covered by Proposition 2.1 and 2.2.

In this section, we consider the wealth process $X^x = (X^x_t)_{t\geqslant 0}$ being an exponential Brownian motion such that

$$X^x_t = x\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right)$$

combined with the value and weighting function of Tversky and Kahneman (1992) as in (1.1) and (1.3). We consider parameters combination such that the drift of the asset is non-negative and the optimal stopping problem is well-posed. A sufficient condition is given by $\frac{\alpha_+}{d_+} < 1 - \frac{2\mu}{\sigma^2} \leqslant 1$ (see Proposition 1.1 in Chapter 1). We study the behaviour of a naive agent with different initial wealth levels x (while keeping his reference point R fixed). The strategy of the naive agent at wealth level x is given by the time zero element of the optimal solution of the problem

$$V(x) = \sup_\tau \mathcal{E}(X^x_\tau - R). \tag{2.11}$$

Problem (2.11) has been solved completely in Chapter 1 and the optimal target prospect $\mathcal{P}^*(x)$ has a (scaled) quantile function in form of (1.11). It contains a point mass on losses set at some level $\ell(x)$ below $\min(x, R)$, a point mass at some level $a(x) > R$ together with a continuous distribution on a semi-infinite interval $(a(x), \infty)$. From Proposition 1.7, we know that the optimal prospect does not allocate any probability mass to the reference level. Hence a naive agent will never stop at the reference level even if he has access to randomised strategies.

However, there exists initial wealth levels x such that x lies inside the support of the optimal prospect $\mathcal{P}^*(x)$. In Figure 2.7 we plot the location $\ell(x)$ of the point mass on losses and the lower limit $a(x)$ of the support of the distribution on gains as functions of the initial wealth x. For small initial wealths $x < x^* = 1.015$, the optimal prospect consists of a point mass on losses at $\ell(x)$ where $\ell(x) < \min\{x, R\}$, and a distribution on gains supported on $[a(x), \infty)$ where $a(x) > \max\{x, R\}$. But for larger initial wealths $x > x^*$, the form of the optimal prospect changes. There is no mass assigned to losses, and moreover $a(x) < x$. Then the initial wealth lies within the support of the optimal prospect.

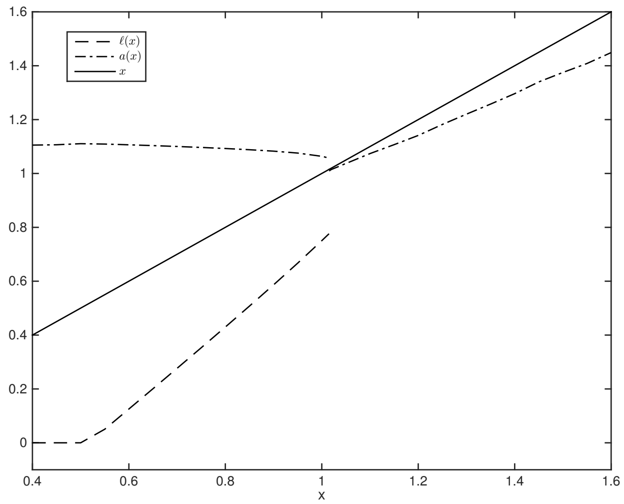


Figure 2.7: The plot of the levels of losses and the lower bounds of gains against different initial wealth levels x. The model takes a martingale exponential Brownian motion and Tversky and Kahneman (1992) value and weighting functions. For $x_0 > x^* = 1.015$, the optimal prospect does not place any probability mass on losses, and the lower bound on the support of the optimal prospect lies below the initial wealth level x. Parameters used are $\alpha_+ = 0.5$, $\alpha_- = 0.9$, $k = 1.25$, $\delta_\pm = 0.7$ and $R = 1$.

The implications for the naive agent in this example are as follows. If at time t the current wealth X_t is such that $X_t < x^*$ then it is not optimal to stop. If the current wealth lies above x^* then it lies within the interval to which the optimal prospect assigns mass, and a naive agent with access to randomised strategies may stop there. Although for a given initial wealth there is no unique optimal stopping rule for the agent with a precommitment device, there is an unique optimal stopping rule within the class of time-homogeneous, Markovian strategies.

At an initial wealth level of x, suppose the precommitting agent plans to attain the optimal prospect $\mathcal{P}^*(x)$ by stopping at y at rate $\Psi(y; x)$ per unit time for some function Ψ. If the naive PT agent adopts strategies from this family, in particular, if he uses the instantaneously-optimal (time-homogeneous, Markovian) strategy of an agent with a precommitment device, then the naive agent stops at x at rate $\Psi(x; x)$. The resultant stopping rate $\Psi(x;x)$ adopted by the naive agent is non-zero for all x above x^*. In this case the realised prospect of the naive agent consists of a density on (x^*, ∞), together with a point mass at 0 (its size is equal to the probability that X never reaches x^*). In other words, a naive agent who can randomise may stop anywhere above x^*.

2.7 Conclusion

This chapter derives new results on prospect theory in a dynamic context. Our first contribution is to observe that, unlike in the expected utility paradigm, PT agents can benefit from following randomised strategies. We investigate the impact of randomised strategies in recombining binomial tree models in the spirit of Barberis (2012) and show that the ability to follow randomised strategies can lead to improvements in PT value.

Our second contribution is to revisit the continuous time model of Ebert and Strack (2015) under an assumption that agents have access to randomised strategies. Ebert and Strack (2015) argue that for any reasonable specification of PT preferences, a naive agent prefers to stop on the first exit from some interval to stopping immediately. We show in a general setting that there is a mixed strategy which is preferred to this first exit strategy, and our mixed strategy may involve stopping at the reference level. Moreover, we provide examples where the optimal prospect for an agent includes mass (in form of an atom or a density) at the initial wealth level. If the naive agent is able to follow randomised strategies, then the agent may realise this optimal prospect with a strategy which involves sometimes stopping. Ebert and Strack (2015) show that under pure strategies, PT preferences and naïvité lead to the "*unrealistic*" conclusion of never stopping. The authors discuss options to evade their never stopping result including dispensing with the probability weighting element of PT or the notion of naïvité. Our results show that a third possibility is to retain probability weighting and naïvité but to allow the agent to use randomised strategies. If the setup is expanded to allow for randomised strategies then the predictions of a dynamic prospect theory model are closer to reality in the sense that they include voluntary cessation of gambling for a naive agent.

Appendix to Chapter 2

2.A Simple sufficient conditions for (2.6).

A pair of conditions which together are sufficient for (2.6) is

$$\operatorname*{argmax}_{p\in[0,1]} \{w_+(p) - p\} + \arg\min_{p\in[0,1]} \left\{ \frac{w_-(p)}{p} \right\} < 1 \tag{2.12}$$

and

$$K > \left(\min_{p\in[0,1]} \frac{w_-(p)}{p} \right)^{-1}. \tag{2.13}$$

Proposition 2.5. *Suppose $w_\pm = w$ where w is a differentiable, inverse-S shaped probability weighting function which is concave on $[0,Q]$ and convex on $[Q,1]$. Let $A(p) = w(p) + w(1-p)$. If A is strictly decreasing on $[0,1/2)$, then $Q \leq 1/2$ and (2.12) holds.*

Proof. Let $q_+ = \operatorname{argmax}_{p\in[0,1]}(w(p) - p)$ and $q_- = \operatorname{argmin}_{p\in[0,1]} \left(\frac{w(p)}{p}\right)$. Since A is strictly decreasing on $[0,1/2)$ it follows that $w'(p) - w'(1-p) < 0$ there. Suppose $Q > \frac{1}{2}$. Then $1 - Q < 1/2$ and $w'(1-Q) < w'(Q)$. But from the shape of w, w' is minimised at Q, which yields a contradiction. Hence $Q \leqslant \frac{1}{2}$.

Define $\overline{w}(p) := 1 - w(1-p)$ which is an inverse-S shaped probability weighting function made by rotating w. Since q_- is a contact point of the largest convex function dominated by w, by symmetry $1 - q_-$ is a contact point of the concave majorant to $\overline{w}$.

Let $q^* \in [0, 1-Q]$ be the solution to the equation $\overline{w}'(q^*) = 1$. As $w'(p) \leqslant w'(1-p) = \overline{w}'(p)$ on $[0,Q]$ and $w'(q_+) = 1 = \overline{w}'(q^*)$, we have $\overline{w}'(q^*) = w'(q_+) \leqslant \overline{w}'(q_+)$. Since $q_+ \leqslant Q \leqslant 1 - Q$ and $\overline{w}'$ is decreasing on $[0, 1-Q]$, this gives $q_+ \leqslant q^*$. Also $\overline{w}'(1-q_-) < 1 = \overline{w}'(q^*)$ and in turn $q^* < 1 - q_-$. It follows that $q_+ \leqslant q^* < 1 - q_-$ or equivalently $q_+ + q_- < 1$. Hence (2.12) holds. □

Corollary 2.6. *If $w_\pm = w$ then for the Tversky and Kahneman (1992) and Goldstein and Einhorn (1987) weighting functions (2.12) holds. Then if K is sufficiently large such that (2.13) holds, (2.6) holds also.*

Proof. The proof of this result makes use of the fact that these weighting functions are inverse-S shaped, and therefore all that needs to be checked is that A is strictly decreasing on $[0,1/2)$, which is a simple exercise in calculus. □

2.B Solutions in the examples of Section 2.5

Define

$$\begin{aligned}
D(\mu) &= \sup_{\int_0^\infty \bar{F}_X(x)dx=\mu=\int_{-\infty}^0 F_X(x)dx} \left\{ \int_0^\infty w_+(\bar{F}_X(x))U'(x)dx - \int_{-\infty}^0 w_-(F_X(x))U'(x)dx \right\} \\
&= \sup_{\int_0^\infty \bar{F}_X(x)dx=\mu=\int_{-\infty}^0 F_X(x)dx} \left\{ \int_0^\infty w_+(\bar{F}_X(x))U'(x)dx - K\int_{-\infty}^0 w_-(F_X(x))dx \right\}
\end{aligned}$$

since $U(x) = Kx$ on $x < 0$ for both specifications. Then $V = \sup_\mu D(\mu)$. The expression for $D(\mu)$ can be rewritten as

$$D(\mu) = \sup_{(h_+,h_-)\in\mathcal{A}_1^2(\mu)} \left\{ \int_0^\infty w_+(h_+(x))U'(x)dx - K\int_0^\infty w_-(h_-(x))dx \right\}$$

where the set $\mathcal{A}_1^2(\mu)$ is given by

$$\mathcal{A}_\phi^2(\mu) = \Big\{ (h_+, h_-) : [0,\infty)^2 \to [0,\infty)^2, h_\pm \text{ decreasing and right continuous}, h_\pm(0) \le 1, \int_0^\infty h_\pm(x)dx = \mu, h_+(0) + h_-(0) \le \phi \Big\}.$$

Then $D(\mu) \le \tilde{D}(\mu)$ where

$$\tilde{D}(\mu) = \sup_{(h_+,h_-)\in\mathcal{A}_2^2(\mu)} \left\{ \int_0^\infty w_+(h_+(x))U'(x)dx - K\int_0^\infty w_-(h_-(x))dx \right\}.$$

In calculating $D(\mu)$ we require the total mass of the target law X to be less than or equal to one (with the understanding that X can be made into a random variable with unit total mass by including an atom at zero). In calculating $\tilde{D}(\mu)$ we make no such requirement. The advantage of considering $\tilde{D}(\mu)$ is that the problems over gains and losses decouple:

$$\tilde{D}(\mu) = \sup_{h\in\mathcal{A}_1^1(\mu)} \left\{ \int_0^\infty w_+(h(x))U'(x)dx \right\} - K \inf_{h\in\mathcal{A}_1^1(\mu)} \left\{ \int_0^\infty w_-(h(x))dx \right\}$$

where

$$\mathcal{A}_\phi^1(\mu) = \{h : [0,\infty) \to [0,\infty), h \text{ decreasing and right continuous}, h(0) \le \phi, \int_0^\infty h(x)dx = \mu\}.$$

Set

$$\tilde{G}(w;\mu) = \sup_{h\in\mathcal{A}_1^1(\mu)} \int_0^\infty w(h(x))U'(x)dx, \qquad \tilde{L}(w;\mu) = \inf_{h\in\mathcal{A}_1^1(\mu)} \int_0^\infty w(h(x))dx.$$

Then

$$\tilde{D}(\mu) = \tilde{G}(w_+;\mu) - K\tilde{L}(w_-;\mu).$$

Our plan is to solve for $\tilde{G}(w_+;\mu)$ and $\tilde{L}(w_-;\mu)$ (and to find respective optimisers h_+^μ, h_-^μ) and hence to find the μ ($\hat{\mu}$ say) which maximises $\tilde{D}(\mu)$. There are two possibilities. If

$h_+^{\hat{\mu}}(0)+h_-^{\hat{\mu}}(0)>1$ then the optimiser for $\tilde{D}=\sup_\mu \tilde{D}(\mu)$ is not feasible for D (and then the optimiser for D depends on a complicated interplay between $h_+(0)$ and $h_-(0)$); alternatively if $h_+^{\hat{\mu}}(0)+h_-^{\hat{\mu}}(0)\leq 1$ then the optimiser for $\tilde{D}=\sup_\mu \tilde{D}(\mu)$ is feasible for D and hence is an optimiser for D. The corresponding optimal target law has $\mathbb{P}(X>x)=h_+^{\hat{\mu}}(x)$ for $x>0$ and $\mathbb{P}(X<x)=h_-^{\hat{\mu}}(-x)$ for $x<0$. If $h_+^{\hat{\mu}}(0)+h_-^{\hat{\mu}}(0)<1$ then the optimal target law includes an atom of size $1-(h_+^{\hat{\mu}}(0)+h_-^{\hat{\mu}}(0))$ at zero.

We first look at the problem for losses which is common to both specifications. Recall $\eta=\min\left(\frac{w_-(p)}{p}\right)$ and $q_-=\operatorname{argmin}\left(\frac{w_-(p)}{p}\right)$. Then $w_-(h)\geq \eta h$ with equality at zero and q_-. For $h\in\mathcal{A}_1^1(\mu)$ we have

$$\int_0^\infty w_-(h(x))dx \geq \int_0^\infty \eta h(x)dx=\eta\mu.$$

Set $h^*(x)=q_- I_{(x<\frac{\mu}{q_-})}$. Then h^* is admissible and $\int_0^\infty w_-(h^*(x))dx=\frac{\mu}{q_-}w_-(q_-)=\eta\mu$. Then h^* is optimal and $\tilde{L}(w_-;\mu)=\eta\mu$.

We complete the solution construction for each specification in the following two subsections.

2.B.1 Specification I

In this case we have $\tilde{G}(w_+;\mu)=\sup_{h\in\mathcal{A}_1^1(\mu)}\int_0^1 w_+(h(x))dx$. Recall that w_+ is a general inverse-S shaped probability weighting function. Let Q_+ be the inflexion point of w_+.

Solution of the problem for gains

Extend the definition of w_+ to $[0,\infty)$ by setting $w_+(p)=1$ for $p>1$. Let W be the smallest concave majorant of w_+. Then $W=w_+$ on $[0,\varphi]$ and W is linear on $[\varphi,1]$ for some $\varphi<Q_+$. Let $\tilde{W}(\lambda)=\sup_{p\in(0,1)}\{W(p)-\lambda p\}$ be the convex dual of W.

For $\lambda>0$, $W(p)\leq\tilde{W}(\lambda)+\lambda p$ and

$$\int_0^1 w_+(h(x))dx \leq \int_0^1 W(h(x))dx \leq \int_0^1 \{\tilde{W}(\lambda)+\lambda h(x)\}dx \leq \tilde{W}(\lambda)+\lambda\mu$$

where we use $\int_0^1 h(x)dx\leq\int_0^\infty h(x)dx=\mu$. Since λ is arbitrary, and W is concave, so that $\inf_\lambda\{\tilde{W}(\lambda)+\lambda\mu\}=W(\mu)$, we have $\int_0^1 w_+(h(x))dx\leq W(\mu)$.

Now, we exhibit an admissible h such that $\int_0^1 w_+(h(x))dx=W(\mu)$. This h is then optimal and $\tilde{G}(w_+;\mu)=W(\mu)$. First consider the degenerate case when $\mu\geq 1$. Take $h(x)=I_{(x<\mu)}$; then $\int_0^1 w_+(h(x))dx=1=W(\mu)$. Now consider the case where $\mu\leq\varphi$. Take $h(x)=\mu I_{(x<1)}$; then $\int_0^1 w_+(h(x))dx=w_+(\mu)=W(\mu)$. Finally consider the case $\varphi<\mu<1$. Take $h(x)=I_{\left(x<\frac{\mu-\varphi}{1-\varphi}\right)}+\varphi I_{\left(\frac{\mu-\varphi}{1-\varphi}\leq x<1\right)}$. Then $\int_0^1 h(x)dx=\frac{\mu-\varphi}{1-\varphi}+\varphi\frac{1-\mu}{1-\varphi}=\mu$.

Moreover, since h takes values 1 and φ, and since $W = w_+$ at these points, and W is linear on $[\varphi, 1]$,

$$\int_0^1 w_+(h(x))dx = w_+(1)\frac{\mu - \varphi}{1-\varphi} + w_+(\varphi)\frac{1-\mu}{1-\varphi} = W(\mu).$$

Solution of the problem for gains and losses

By definition

$$\tilde{D} = \sup_\mu\{\tilde{G}(w_+;\mu) - K\tilde{L}(w_-;\mu)\} = \sup_\mu\{W(\mu) - K\eta\mu\} = \tilde{W}(K\eta).$$

Under (2.6) and more especially the fact that $q_+ := \operatorname{argmax}_{p\in[0,1]}\{w_+(p) - K\eta p\} < 1$ we have that the optimal μ given by $\nu = q_+$ satisfying $0 \le \nu < \varphi$. For this ν, the solution of the problem for gains is a mass of size q_+ at 1, and the solution of the problem for losses is a mass of size q_- at $-q_+/q_-$. Moreover, under (2.6) we have $q_+ + q_- < 1$. Then $\tilde{D} = D$ and the optimiser for the problem is the prospect $\mathcal{P}^* = (-q_+/q_-, q_-; 0, 1 - q_+ - q_-; 1, q_+)$ which agrees with (2.10).

2.B.2 Specification II

In this case we have $\tilde{G}(w_+;\mu) = \sup_{h\in\mathcal{A}_1^1(\mu)} \int_0^\infty \frac{1}{1+x} w_+(h(x))dx$. Recall that w_+ is piecewise linear.

Solution of the problem for gains

Let W be the smallest concave majorant of w_+. We solve the maximisation problem for the concave probability weighting function W. This gives an upper bound for the problem with probability weighting function w_+. Then we show that this bound can be attained.

Set $\gamma = \frac{1-\alpha q_+}{1-q_+}$. Then $\gamma \in (0,1)$ and

$$W(p) = \begin{cases} \alpha p, & 0 \le p \le q_+, \\ \alpha q_+ + \gamma(p - q_+), & q_+ < p \le 1. \end{cases}$$

We want to find $\tilde{G}(W,\mu) = \sup_{h\in\mathcal{A}_1^1(\mu)} \hat{G}_W(\mu, h)$ where

$$\hat{G}_W(\mu,h) = \int_0^\infty \frac{1}{1+x}\left(\alpha h(x) I_{\{h(x)\le q_+\}} + (\alpha q_+ + \gamma(h(x) - q_+))I_{\{h(x)>q_+\}}\right) dx.$$

Suppose first that $\mu \le \left(\frac{\alpha}{\gamma} - 1\right) q_+$. Set

$$\mathcal{I}(x) = \max_{g\in[0,1]}\left\{\frac{1}{1+x}\left(\alpha g I_{\{g\le q_+\}} + (\alpha q_+ + \gamma(g - q_+))I_{\{g>q_+\}}\right) - \frac{\alpha q_+ g}{\mu + q_+}\right\}. \tag{2.14}$$

Then, for $h \in \mathcal{A}_1^1(\mu)$,

$$\hat{G}_W(\mu, h) \leq \frac{\alpha q_+ \mu}{\mu + q_+} + \int_0^\infty \mathcal{I}(x)dx.$$

From the piecewise linear structure of the objective function the maximum in (2.14) can only occur at $g \in \{0, q_+, 1\}$ and

$$\mathcal{I}(x) = \max\left\{0, \frac{\alpha q_+}{1+x} - \frac{\alpha q_+^2}{\mu + q_+}, \frac{1}{1+x} - \frac{\alpha q_+}{\mu + q_+}\right\}.$$

Then since

$$\begin{aligned}\left(\frac{\alpha q_+}{1+x} - \frac{\alpha q_+^2}{\mu + q_+}\right) - \left(\frac{1}{1+x} - \frac{\alpha q_+}{\mu + q_+}\right) &= \frac{\alpha q_+}{\mu + q_+}(1 - q_+) - \frac{1 - \alpha q_+}{1+x} \\ &\geq \gamma(1 - q_+) - (1 - \alpha q_+) = 0\end{aligned}$$

we have

$$\mathcal{I}(x) = \begin{cases} 0, & x \geq \frac{\mu}{q_+}, \\ \alpha q_+ \left(\frac{1}{1+x} - \frac{q_+}{\mu + q_+}\right), & x < \frac{\mu}{q_+}. \end{cases}$$

Then for any $h \in \mathcal{A}_1^1(\mu)$

$$\hat{G}_W(\mu, h) \leq \frac{q_+ \alpha \mu}{\mu + q_+} + \int_0^{\mu/q_+} \left[\frac{\alpha q_+}{1+x} - \frac{\alpha q_+^2}{\mu + q_+}\right] dx = \alpha q_+ \ln\left(1 + \frac{\mu}{q_+}\right). \tag{2.15}$$

Further, if $h(x) = q_+ I_{\{x < \mu/q_+\}}$ then $h \in A_1^1(\mu)$ and there is equality in (2.15). Hence $\tilde{G}(W, \mu) = \alpha q_+ \ln\left(1 + \frac{\mu}{q_+}\right)$.

Finally, for the non-negative random variable $X^* = X^*(\mu)$ with $\overline{F}_{X^*}(x) = q_+ I_{\{x < \mu/q_+\}}$ for $x \in (0, \infty)$ we find

$$\alpha q_+ \ln\left(1 + \frac{\mu}{q_+}\right) = \int_0^\infty \frac{1}{1+x} w_+(\overline{F}_{X^*}(x))dx \leq \tilde{G}(w_+, \mu) \leq \tilde{G}(W, \mu) = \alpha q_+ \ln\left(1 + \frac{\mu}{q_+}\right)$$

and hence $\tilde{G}(w_+, \mu) = \alpha q_+ \ln\left(1 + \frac{\mu}{q_+}\right)$.

Now suppose $\mu > \left(\frac{\alpha}{\gamma} - 1\right) q_+$. Set

$$\begin{aligned}\mathcal{J}(x) &= \max_{g \in [0,1]} \left\{\frac{1}{1+x}\left(\alpha g I_{\{g \leq q_+\}} + (\alpha q_+ + \gamma(g - q_+)) I_{\{g > q_+\}}\right) - \frac{g}{1+\mu}\right\} \qquad (2.16) \\ &= \max\left\{0, \frac{\alpha q_+}{1+x} - \frac{q_+}{1+\mu}, \frac{1}{1+x} - \frac{1}{1+\mu}\right\}.\end{aligned}$$

We find

$$\mathcal{J}(x) = \begin{cases} \frac{1}{1+x} - \frac{1}{1+\mu}, & 0 \leq x \leq \gamma(1+\mu) - 1, \\ q_+\left(\frac{\alpha}{1+x} - \frac{1}{1+\mu}\right), & \gamma(1+\mu) - 1 < x \leq \alpha(1+\mu) - 1, \\ 0, & \alpha(1+\mu) - 1 < x, \end{cases}$$

and the optimiser in (2.16) is $g = h^*(x)$ where

$$h^*(x) = \begin{cases} 1, & 0 \le x \le \gamma(1+\mu) - 1, \\ q_+, & \gamma(1+\mu) - 1 < x \le \alpha(1+\mu) - 1, \\ 0, & \alpha(1+\mu) - 1 < x. \end{cases}$$

Then for $h \in \mathcal{A}_1^1(\mu)$ we have

$$\hat{G}_W(\mu, h) \le \frac{\mu}{1+\mu} + \int_0^\infty \mathcal{J}(x)dx = \alpha q_+ \ln\left(\frac{\alpha}{\gamma}\right) + \ln\gamma(1+\mu).$$

As before, if we define X^* via $\overline{F}_{X^*}(x) = h^*(x) \in \mathcal{A}_1^1(\mu)$ then

$$\begin{aligned} \alpha q_+ \ln\left(\frac{\alpha}{\gamma}\right) + \ln\gamma(1+\mu) &= \int_0^\infty \frac{1}{1+x} w_+(\overline{F}_{X^*}(x)) \\ &\le \tilde{G}(w_+, \mu) \le \tilde{G}(W, \mu) \le \alpha q_+ \ln\left(\frac{\alpha}{\gamma}\right) + \ln\gamma(1+\mu). \end{aligned}$$

Hence $\tilde{G}(w_+, \mu) = \alpha q_+ \ln\left(\frac{\alpha}{\gamma}\right) + \ln\gamma(1+\mu)$.

In summary,

$$\tilde{G}(w_+, \mu) = \begin{cases} \alpha q_+ \ln\left(1 + \frac{\mu}{q_+}\right), & 0 < \mu \le q_+\left(\frac{\alpha}{\gamma} - 1\right), \\ \alpha q_+ \ln\left(\frac{\alpha}{\gamma}\right) + \ln\gamma(1+\mu), & \mu > q_+\left(\frac{\alpha}{\gamma} - 1\right). \end{cases}$$

Note that $\tilde{G}(w_+, \cdot)$ is continuously differentiable (and concave) in its second argument. Let $\mathcal{G}$ be the inverse to the derivative of $\tilde{G}$. Then $\mathcal{G}(y) = (\tilde{G}')^{-1}(y) = (\frac{1}{y} - 1)I_{\{y<\gamma\}} + q_+(\frac{\alpha}{y} - 1)I_{\{y \ge \gamma\}}$.

Solution of the problem for gains and losses

By definition $\tilde{D} = \sup_\mu\{\tilde{G}(w_+;\mu) - K\tilde{L}(w_-;\mu)\}$. We find that the supremum over μ is attained at $\nu = \mathcal{G}(\frac{Kw_-(q_-)}{q_-}) = q_+\left(\frac{\alpha q_-}{Kw_-(q_-)} - 1\right)$. Condition (2.9) is equivalent to $\frac{Kw_-(q_-)}{q_-} \in (\gamma, \alpha)$ which ensures $\nu > 0$. For $\mu = \nu < \left(\frac{\alpha}{\gamma} - 1\right) q_+$, the solution of the problem for gains is a mass of size q_+ at $\frac{\nu}{q_+}$, and the solution of the problem for losses is a mass of size q_- at $-\frac{\nu}{q_-}$. Moreover, by hypothesis $q_+ + q_- < 1$. Then $\tilde{D} = D$ and the optimiser for the problem is the prospect $\mathcal{P}^* = (-\frac{\nu}{q_-}, q_-; 0, 1 - q_+ - q_-; \frac{\nu}{q_+}, q_+)$.

2.C Constructions of $w_\pm$ satisfying all the conditions in Propositions 2.1, 2.2 and 2.3.

In this section, we demonstrate it is possible to construct probability weighting functions $w_\pm$ which simultaneously satisfy all the conditions in (2.2) and (2.3) and those listed in

Section 2.5.1 for each specification.

Under Specification I, the $w_\pm$ can be any general inverse-S shaped functions. It follows from the discussion in Section 2.4.2 and Appendix 2.A that, for example, the Tversky and Kahneman (1992) weighting functions with $\delta = \delta_+ = \delta_-$ can satisfy all the required conditions if K is sufficiently large.

Now we consider Specification II. Suppose w_- is continuously differentiable and such that $\Lambda := \sqrt{w'_-(1-)\frac{w_-(q_-)}{q_-}} > 1$. (The Tversky and Kahneman (1992) and Goldstein and Einhorn (1987) weighting functions satisfy this condition for all parameter combinations, as does the piecewise linear weighting function if $\alpha\beta \geq 1$). Fix $\lambda \in (1, \Lambda)$ and suppose $K > 1$ is such that $K \in (\frac{q_-}{w_-(q-)}, \frac{w'_-(1-)}{\lambda^2})$. Then, since $w'_-(1-) > K\lambda^2$, there exists $\overline{q} > 0$ such that $w'_-(1-q) > \lambda K$ for $q \leq \overline{q}$ and then $w_-(1-q) < 1 - \lambda K q < 1 - Kq$ for $q \leq \overline{q}$. Now choose a piecewise linear w_+ with $\alpha \in (K, \lambda K)$, $\beta \in (0,1)$ and $q_+ < \min\{\frac{1-\beta}{2(\alpha-\beta)}, \overline{q}, 1-q_-\}$. Then for $\hat{p} \in (0, q_+)$, $w_+(\hat{p}) = \alpha\hat{p} > K\hat{p} > \frac{K\hat{p}}{1+(K-1)\hat{p}}$ and $w_-(1-\hat{p}) < 1 - K\hat{p} < \frac{1-\hat{p}}{1+(K-1)\hat{p}}$. Thus (2.2) holds. Further, $w'_-(1-\hat{p}) > \lambda K > \alpha = w'_+(\hat{p})$ so that (2.3) holds. Finally we have $q_+ < 1 - q_-$ and $\frac{1-\alpha q_+}{1-q_+} < 1 < \frac{K w_-(q_-)}{q_-} < K < \alpha$ such that the assumptions under Specification II hold.

Chapter 3

A multi-asset investment and consumption problem with transaction costs

"What is so terrible about transaction costs? On what basis are they considered the ultimate evil, so that their minimization must override all other considerations of choice, freedom, and justice?"
— Murray Rothbard, *The Myth of Neutral Taxation*

3.1 Introduction

The study of optimal investment and consumption problem over an infinite time horizon dated back to the seminal work of Merton (1969), in which a CRRA agent can trade in a frictionless market consisting of one risk-free bond and one risky asset with price process following a geometric Brownian motion. Merton provides an explicit solution to the problem: the agent should invest a constant fraction of his wealth into the risky asset and maintain a fixed consumption rate per unit wealth. Despite the mathematical elegance of the solution, it is impractical to implement such an investment strategy involving continuous rebalancing in view of transaction costs in the real world.

Ever since then researchers have been looking to incorporate market frictions in the Merton model. Magill and Constantinides (1976) provide intuitions of the form of the optimal strategy in presence of proportional transaction costs. The agent should trade in a minimal fashion to keep the fraction of wealth allocated to the risky asset within a fixed interval which is referred to as the no-transaction region. This conjectured strategy is characterised mathematically by Davis and Norman (1990) in form of a local time trading rule, and its optimality is proved by a standard Hamilton-Jacobi-Bellman (HJB) verification argument under some technical assumptions. Shreve and Soner (1994) utilise the notion of viscosity solution to overcome some of the technical restrictions in Davis and Norman (1990) and re-establish many of their results. More recently, a dual formulation based on the concept of shadow price has emerged, starting with Kallsen and Muhle-Karbe (2010) in the context of logarithm utility and later extended to power utility by Choi et al. (2013) and Herczegh and Prokaj (2015). Choi et al. (2013) derive the conditions on the model parameters leading to a well-posed problem. Using a primal HJB approach, Hobson et al. (2016) re-prove these well-posedness conditions and along the way provide a number of new analytical results on the comparatives statics of the problem.

The existing literature on this subject has been focusing on a market with one risky asset only, and thus it is a natural direction to generalise the results to a model with multiple risky assets. However, the extension is difficult and the limited progress made so far in a higher dimensional setting is mostly confined to technical characterisations of the value function/trading strategy, numerical studies or asymptotic analysis. Akian et al. (1995) show that the value function is the unique viscosity solution to a HJB variational inequality. In a model with two assets, Chen and Dai (2013) characterise the shape of the no-transaction region. Explicit solutions in a multi-asset problem with transaction costs are rare in general. An exception is the special case studied by Liu (2004) where the risky assets are uncorrelated and the agent has a CARA utility function. Otherwise, one has to resort to numerical methods such as a policy improvement algorithm proposed by Muthuraman and Kumar (2006), or a Markov chain approximation scheme of Collings and Haussmann (1999) which they prove the convergence. Some asymptotic results are available in the cases with small transaction costs. See Possamaï et al. (2015) for example.

In this chapter we consider the problem with a risk-free bond and two risky assets. Transactions in the first risky asset are costless, but transactions in the second risky asset, which we term the illiquid asset, incur proportional costs. (More generally, we may have several risky assets on which no transaction costs are payable. By a mutual fund theorem, this general case can be reduced to the case with a single liquid, risky asset.)

The methodology adopted in this chapter is a primal HJB approach based on Hobson et al. (2016) which considers the classical case with one risky asset only. We postulate that the value function depends on four variables: the total amount of liquid wealth, the price level of the illiquid asset, number of units of the illiquid asset held and time. Although the second asset in our model leads to a much more complicated HJB equation, the transformation scheme used in Hobson et al. (2016) can indeed be employed in our current setting and we obtain a reduced problem with a similar structure. The main difficulty lies with the extra parameters entering the problem which make the analysis of the well-posedness conditions and comparative statics a much more challenging task. But through a detailed study on the underlying differential equations, we are able to extend the ideas from the one-dimensional case in Hobson et al. (2016) to the more general setting considered in this chapter. This in turn allows us to make significant progress towards understanding the analytical behaviours of the solution in a multi-asset setup.

Our first achievement is to show that the problem of finding the free boundaries associated with the HJB variational inequality and the value function can be reduced to the study of a boundary crossing problem for a family of solutions to a class of first order ordinary differential equations parametrised by the initial values. This allows us to characterise precisely the parameter combinations for which the problem is well-posed (Theorem 3.4), and in those cases to give an expression for the value function (Theorem 3.5). These results extend Choi et al. (2013) and Hobson et al. (2016) to the case of multiple risky assets.

Our second achievement is to make definitive statements about the comparative statics for the problem. We focus on the boundaries of the no-transaction wedge and the certainty equivalent value of the holdings in the illiquid asset. Amongst other results, we prove (see Theorem 3.9 and Corollary 3.10 for precise statements) that as the drift on the illiquid asset improves, the agent aims to keep a larger fraction of his total wealth in the illiquid asset, in the sense that the critical ratios at which sales and purchases take place are increasing in the drift. Conversely, as the agent becomes more impatient, the agent keeps a smaller fraction of wealth in the illiquid asset. Further, we prove (Theorem 3.11 and Corollary 3.12) that as the drift on the illiquid asset improves, or as the agent becomes less impatient, the certainty equivalent value of the holdings in the illiquid asset increases. See Section 3.6 for a more detailed discussion.

The remainder of the chapter is as follows. In the next section we formulate the problem. In Section 3.3 we derive the HJB equation and give heuristics showing how it can be converted to a free boundary value problem involving a first order differential equation. Then we can state our main results in Section 3.4 on the existence of a solution. In Section 3.5 we discuss the various cases which arise. In Section 3.6 we discuss the comparative statics of the problem, before Section 3.7 concludes. Materials on the solution of the free boundary value problem, the verification argument for the HJB equation, and other lemmas on the analysis of solutions of the differential equations are relegated to an appendix.

3.2 The problem

The economy consists of one money market instrument paying constant interest rate $r > 0$ and two risky assets, one of which is liquidly traded while the other one is illiquid. There are no transaction costs associated with trading in the liquid asset. Meanwhile, trading in the illiquid asset incurs a proportional transaction cost $\lambda \in [0, \infty)$ on purchases and $\gamma \in [0, 1)$ on sales where not both λ and γ are zero. Let $(S, Y) = (S_t, Y_t)_{t \geqslant 0}$ be the price processes of the liquid and illiquid assets respectively. The price dynamics are given by

$$(S_t, Y_t) = \left(S_0 \exp\left((\mu - \frac{\sigma^2}{2})t + \sigma B_t \right), Y_0 \exp\left((\alpha - \frac{\eta^2}{2})t + \eta W_t \right) \right)$$

where (B, W) is a pair of Brownian motions with correlation coefficient $\rho \in (-1, 1)$. Write $\beta := (\mu - r)/\sigma$ and $\nu := (\alpha - r)/\eta$ for the Sharpe ratio of the liquid and illiquid asset respectively.

Let Θ_t be the number of units of the illiquid asset held by an agent at time t. Then $\Theta_t = \Theta_0 + \Phi_t - \Psi_t$ where $\Phi = (\Phi_t)_{t \geqslant 0}$ and $\Psi = (\Psi_t)_{t \geqslant 0}$ are both increasing, non-negative processes representing the cumulative units of purchases and sales respectively of the illiquid asset. Let $C = (C_t)_{t \geqslant 0}$ be the non-negative consumption rate process of the agent and $\Pi = (\Pi_t)_{t \geqslant 0}$ be the cash value of holdings in the risky liquid asset. We assume Θ, C and Π are progressively measurable and right-continuous. If $X = (X_t)_{t \geqslant 0}$ is the total value of the liquid instruments (cash and the liquid risky asset) then, assuming transaction costs are paid in cash,

$$\begin{aligned} dX_t &= r(X_t - \Pi_t)dt + \frac{\Pi_t}{S_t} dS_t - C_t dt - Y_t(1 + \lambda)d\Phi_t + Y_t(1 - \gamma)d\Psi_t \\ &= [(\mu - r)\Pi_t + rX_t - C_t]\, dt - Y_t(1 + \lambda)d\Phi_t + Y_t(1 - \gamma)d\Psi_t + \sigma \Pi_t dB_t. \end{aligned}$$

We say that a portfolio (X, Θ) is solvent at time t if its instantaneous liquidation value is non-negative, that is

$$X_t + \Theta_t^+ Y_t(1 - \gamma) - \Theta_t^- Y_t(1 + \lambda) \geqslant 0.$$

A consumption/investment strategy (C, Π, Θ) is said to be admissible if the resulting portfolio is solvent at the current time and at all the future time points. Write $\mathcal{A}(t, x, y, \theta)$ for the set of admissible strategies with initial time-t value $(X_{t-} = x, Y_t = y, \Theta_{t-} = \theta)$.

We assume the agent has a CRRA utility function with risk aversion parameter $R \in (0,\infty) \setminus \{1\}$. His objective is to find an optimal strategy which maximises the expected lifetime discounted utility from consumption. The problem is thus to find

$$V(x, y, \theta) = \sup_{(C,\Pi,\Theta)\in\mathcal{A}(0,x,y,\theta)} \mathbb{E}\left(\int_0^\infty e^{-\delta s}\frac{C_s^{1-R}}{1-R}ds\right) \tag{3.1}$$

where δ is the agent's subjective discount rate.

We will call $X_t + \Theta_t Y_t$ the paper wealth of the agent. In our parametrisation a key quantity will be $P_t := \frac{\Theta_t Y_t}{X_t+\Theta_t Y_t}$, the proportion of paper wealth invested in the illiquid asset.

3.3 The HJB equation and a free boundary value problem

3.3.1 Deriving the HJB equation

Let

$$\mathcal{V}(x, y, \theta, t) = \sup_{(C,\Pi,\Theta)\in\mathcal{A}(t,x,y,\theta)} \mathbb{E}\left(\int_t^\infty e^{-\delta s}\frac{C_s^{1-R}}{1-R}ds\right)$$

be the forward-starting value function from time t. Inspired by the analysis in the classical case involving a single risky asset only, we postulate that the value function has the form

$$\mathcal{V}(x, y, \theta, t) = e^{-\delta t}V(x, y, \theta) = \Upsilon\frac{e^{-\delta t}(x + y\theta)^{1-R}}{1-R}G\left(\frac{y\theta}{x+y\theta}\right) \tag{3.2}$$

for some strictly positive function G to be determined and Υ a convenient scaling constant which will help simplify the HJB equation. We take $\Upsilon = \left(\frac{b_1}{Rb_4}\right)^{-R}$ where b_1 and b_4 are constants to be defined below in Section 3.3.2 in terms of the financial parameters associated with the underlying problem. For the present we assume that G is smooth and use heuristic arguments to derive a characterisation of the candidate value function. Later we will outline a verification argument that this candidate value function coincides with the solution of the corresponding optimal investment/consumption problem, and therefore deduce the necessary smoothness properties of $\mathcal{V}$ and G.

Building on the intuition developed by Magill and Constantinides (1976) and Davis and Norman (1990) we expect that the optimal strategy of the agent is to trade the illiquid asset only when P_t falls outside a certain interval $[p_*, p^*]$ to be identified. Due to the solvency restriction, we must have $-\frac{1}{\lambda} \leqslant P_t \leqslant \frac{1}{\gamma}$ and $[p_*, p^*] \subseteq [-\frac{1}{\lambda}, \frac{1}{\gamma}]$. Whenever $P_t < p_*$, the agent purchases the illiquid asset to bring P_t back to p_*. Hence for an initial position

(x,θ) such that $p=\frac{y\theta}{x+y\theta}<p_*$, the number of units of illiquid asset to be purchased is given by $\phi=\frac{xp_*-(1-p_*)y\theta}{y(1+\lambda p_*)}$ such that $\frac{y(\theta+\phi)}{x+y(\theta+\phi)-y(1+\lambda)\phi}=p_*$. The value function does not change on this transaction, and hence we deduce that for $-\frac{1}{\lambda}\leqslant p<p_*$,

$$(x+y\theta)^{1-R}G(p)=[x+y(\theta+\phi)-y(1+\lambda)\phi]^{1-R}G(p_*)$$

and in turn

$$G(p)=\left(\frac{1+\lambda p}{1+\lambda p_*}\right)^{1-R}G(p_*)=A_*(1+\lambda p)^{1-R} \tag{3.3}$$

where $A_*:=G(p_*)(1+\lambda p_*)^{R-1}$. Similar consideration leads to the conclusion that

$$G(p)=\left(\frac{1-\gamma p}{1-\gamma p^*}\right)^{1-R}G(p^*)=A^*(1-\gamma p)^{1-R} \tag{3.4}$$

for $p^*<p\leqslant\frac{1}{\gamma}$ where $A^*:=(1-\gamma p^*)^{R-1}G(p^*)$.

Consider $M=(M_t)_{t\geqslant 0}$ defined via

$$M_t:=\int_0^t e^{-\delta s}\frac{C_s^{1-R}}{1-R}ds+e^{-\delta t}V(X_t,Y_t,\Theta_t).$$

We expect M to be a supermartingale in general, and a martingale under the optimal strategy. Suppose V is $C^{2\times2\times1}$. Then applying Ito's lemma we find

$$\begin{aligned}
&e^{\delta t}dM_t\\
&=\frac{C_t^{1-R}}{1-R}dt+V_xdX_t+\frac{1}{2}V_{xx}d[X]_t+V_ydY_t+\frac{1}{2}V_{yy}d[Y]_t+V_\theta d\Theta_t+V_{xy}d[X,Y]_t-\delta V dt\\
&=\left(\frac{C_t^{1-R}}{1-R}-V_xC_t+\frac{\sigma^2}{2}V_{xx}\Pi_t^2+((\mu-r)V_x+\sigma\eta\rho V_{xy}Y_t)\Pi_t+rV_xX_t+\alpha V_yY_t+\frac{\eta^2}{2}V_{yy}Y_t^2-\delta V\right)dt\\
&\qquad+(V_\theta-(1+\lambda)V_xY_t)d\Phi_t+(V_xY_t(1-\gamma)-V_\theta)d\Psi_t+\sigma V_x\Pi_t dB_t+\eta V_yY_tdW_t.
\end{aligned}$$

Further assume V is strictly increasing and concave in x. Then on maximising the drift term with respect to C_t and Π_t and setting the resulting maxima to zero, we obtain the HJB equation over the no-transaction region:

$$\frac{R}{1-R}V_x^{1-1/R}+rxV_x+\alpha yV_y+\frac{\eta^2}{2}y^2V_{yy}-\frac{(\beta V_x+\eta\rho yV_{xy})^2}{2V_{xx}}-\delta V=0. \tag{3.5}$$

3.3.2 Reduction to a first order free boundary value problem

Define the auxiliary parameters b_1, b_2, b_3 and b_4 as

$$b_1=\frac{2\left[\delta-r(1-R)-\frac{\beta^2(1-R)}{2R}\right]}{\eta^2(1-\rho^2)},\quad b_2=\frac{\beta^2-2R\eta\rho\beta+\eta^2R^2}{\eta^2R^2(1-\rho^2)},\quad b_3=\frac{2(\nu-\beta\rho)}{\eta(1-\rho^2)},$$
$$b_4=\frac{2}{\eta^2(1-\rho^2)}.$$

It will turn out that the optimal investment and consumption problem depends on the original parameters only through these auxiliary parameters and the risk aversion level R.

Here b_1 plays the role of a 'normalised discount factor', which adjusts the discount factor to allow for numeraire growth effects and for investment opportunities in the liquid risky asset. b_4 is related to the level of the 'idiosyncratic volatility' of the illiquid asset. The parameter b_3 is the 'effective Sharpe ratio, per unit of idiosyncratic volatility' of the illiquid asset. The parameter b_2 is the hardest to interpret: essentially it is a nonlinearity factor which arises from the multi-dimensional structure of the problem. Note that $b_2 = 1 + \frac{1}{1-\rho^2}\left(\frac{\beta}{\eta R} - \rho\right)^2 \geqslant 1$.

In the sequel we will work with the following assumption.

Standing Assumption 3.1. *Throughout the chapter we assume $b_1 > 0$, $b_2 > 1$ and $b_3 > 0$.*

The rationale for imposing $b_1 > 0$ is that $b_1 > 0$ is necessary to ensure well-posedness of the Merton problem in the absence of the illiquid asset. (If $R < 1$ and $b_1 \leqslant 0$, the value function is infinite for the Merton problem. Conversely, if $R > 1$ and $b_1 \leqslant 0$, then for every admissible strategy the expected discounted utility of consumption equals $-\infty$). In contrast, the assumption $b_3 > 0$ is not necessary. However, the advantage of working with a positive effective Sharpe ratio of the illiquid asset ($b_3 > 0$) is that the no-transaction wedge is contained in the first two quadrants of the $(x, y\theta)$ plane. The assumption $b_3 > 0$ reduces the number of cases to be considered in our analysis, and facilitates the clarity of the exposition, but the methods and results developed in this chapter can be extended easily to the case of an illiquid asset with non-positive effective Sharpe ratio.

The case $b_2 = 1$ is rather special and we exclude it from our analysis. One scenario in which we naturally find $b_2 = 1$ is if $\beta = 0 = \rho$. In this case there is neither a hedging motive, nor an investment motive for holding the liquid risky asset. Essentially then, the investor can ignore the presence of the liquid risky asset, reducing the dimensionality of the problem. This problem is the subject of Hobson et al. (2016). If $b_2 = 1$ then the solution n we define in the next paragraph may pass through singular points. See Choi et al. (2013) or Hobson et al. (2016) for a discussion of some of the issues.

We adopt the same transformation as Hobson et al. (2016) to reduce the order of the HJB equation. Recall the relationship between $\mathcal{V}$ and G in (3.2) and the definition $p = \frac{y\theta}{x+y\theta}$. Away from $p = 1$, set $h(p) = \text{sgn}(1-p)|1-p|^{R-1}G(p)$, $w(h) = p(1-p)\frac{dh}{dp}$, $W(h) = \frac{w(h)}{(1-R)h}$, let $N = W^{-1}$ be the inverse function to W and set $n(q) = |N(q)|^{-1/R}|1-q|^{1-1/R}$. Then, we show in Appendix 3.A that (3.5) can be transformed into a first order differential equation

$$n'(q) = O(q, n(q)) \tag{3.6}$$

where

$$O(q,n) = \frac{(1-R)n}{R(1-q)} - \frac{2(1-R)^2 qn/R}{2(1-R)(1-q)\left[(1-R)q+R\right] - \varphi(q,n) - \operatorname{sgn}(1-R)\sqrt{\varphi(q,n)^2 + E(q)^2}} \tag{3.7}$$

with

$$\begin{aligned}\varphi(q,n) &:= b_1(n-1) + (1-R)(b_3 - 2R)q + (2-b_2)R(1-R),\\ E(q)^2 &:= 4R^2(1-R)^2(b_2-1)(1-q)^2.\end{aligned}$$

Define the quadratic

$$m(q) := \frac{R(1-R)}{b_1}q^2 - \frac{b_3(1-R)}{b_1}q + 1 \tag{3.8}$$

and the algebraic function

$$\ell(q) := m(q) + \frac{1-R}{b_1}q(1-q) + \frac{(b_2-1)R(1-R)}{b_1}\frac{q}{(1-R)q+R}. \tag{3.9}$$

Note that m has a turning point (a minima if $R < 1$ and a maxima if $R > 1$) at $q_M := \frac{b_3}{2R}$ and set $m_M := m(q_M) = 1 - \frac{b_3^2(1-R)}{4b_1 R}$.

The following are the key properties of the function O. They are special cases of a more complete set of properties given in Lemma 3.6 below.

Lemma 3.2. *1. $O(q,n)$ can be extended to $q = 1$ by continuity on $(1-R)n < (1-R)\ell(1)$;*

2. On $(0,\infty)\times(0,\infty)$, $O(q,n) = 0$ if and only if $n = m(q)$;

3. For given R and q the sign of $O(q,n)$ depends only on the signs of $n - m(q)$ and $\ell(q) - n$.

Now we apply the same transformations which took (3.5) to (3.6) to the value function on the purchase and sale regime. For $-\frac{1}{\lambda} \leqslant p < p_*$, $G(p) = A_*\left(1+\lambda p\right)^{1-R}$ as given by (3.3). Then

$$w(h) = p(1-p)\frac{dh}{dp} = p(1-p)(1-R)h\left[\frac{\lambda}{1+\lambda p} + \frac{1}{1-p}\right] = (1-R)h\left[\frac{p(1+\lambda)}{1+\lambda p}\right]$$

and $|1 - W(h)| = \frac{|1-p|}{1+\lambda p} = \left(\frac{A_*}{|h|}\right)^{1/(1-R)}$. It follows that $n(q) = (A_*)^{-1/R}$. This expression holds for $-\frac{1}{\lambda} \leqslant p < p_*$ on which $q = W(h) = \frac{(1+\lambda)p}{1+\lambda p}$. The equivalent range in q is thus given by $q < q_* := \frac{(1+\lambda)p_*}{1+\lambda p_*}$. Similarly on the sale region we have $n(q) = (A^*)^{-1/R}$ for $q > q^* := \frac{(1-\gamma)p^*}{1-\gamma p^*}$.

The $C^{2\times2\times1\times1}$ smoothness of the original value function $\mathcal{V}$ now translates into C^1 smoothness of the transformed value function n. Hence we are looking for a continuously differentiable function n and boundary points (q_*, q^*) solving (3.6) on $q \in (q_*, q^*)$ with $n(q) = (A_*)^{-1/R}$ for $q \leq q_*$ and $n(q) = (A^*)^{-1/R}$ for $q \geq q^*$. First order smoothness of n at the boundary points forces $n'(q_*) = n'(q^*) = 0$. By Lemma 3.2, $n'(q) = O(q, n(q)) = 0$ if and only if $n(q) = m(q)$. Hence the free boundary points must be given by the q-coordinates where n intersects the quadratic m. The free boundary value problem now becomes solving $n'(q) = O(q, n(q))$ on $q \in (q_*, q^*)$ subject to $n(q_*) = m(q_*)$ and $n(q^*) = m(q^*)$.

As an example, suppose $R < 1$ and $m_M > 0$. Fix $u \in (0, q_M)$. As we will show later in Section 3.5, the solution to (3.6) with initial value $(u, m(u))$ is decreasing. We are interested in when this solution crosses m again, and we call this point $\zeta(u)$. Then we have a family of solutions $(n_u(q))_{u\leq q\leq\zeta(u)}$ to (3.6) with $n(u) = m(u)$ and $n(\zeta(u)) = m(\zeta(u))$. The solution we want is the one which is consistent with the given transaction costs. Our approach is based on the same idea as in Hobson et al. (2016). Let $\xi = \frac{\lambda+\gamma}{1-\gamma} > 0$ be the round-trip transaction cost. Suppose for now $1 \notin [p_*, p^*]$ and in turn $1 \notin [q_*, q^*]$. Exploiting the relationships that $q_* = \frac{(1+\lambda)p_*}{1+\lambda p_*}$ and $q^* = \frac{(1-\gamma)p^*}{1-\gamma p^*}$, we have

$$\ln(1+\xi) = \ln(1+\lambda) - \ln(1-\gamma) = \int_{p_*}^{p^*} \frac{dp}{p(1-p)} - \int_{q_*}^{q^*} \frac{dq}{q(1-q)}.$$

Then, using the definitions of w, N and O,

$$\begin{aligned}
\ln(1+\xi) &= \int_{h_*}^{h^*} \frac{dh}{w(h)} - \int_{q_*}^{q^*} \frac{dq}{q(1-q)} \\
&= \int_{q_*}^{q^*} \frac{N'(q)dq}{(1-R)qN(q)} - \int_{q_*}^{q^*} \frac{dq}{q(1-q)} \\
&= \int_{q_*}^{q^*} \frac{R}{q(1-R)} \left(\frac{N'(q)}{RN(q)} - \frac{1-R}{R(1-q)} \right) dq \\
&= \int_{q_*}^{q^*} \left(-\frac{R}{q(1-R)} \frac{O(q,n(q))}{n(q)} \right) dq \qquad (3.10)
\end{aligned}$$

where in the last line we use the fact that $\frac{O(q,n(q))}{n(q)} = \frac{n'(q)}{n(q)} = \frac{1-R}{R(1-q)} - \frac{1}{R}\frac{N'(q)}{N(q)}$. Hence the required solution from the free boundary value problem is the one such that

$$\ln(1+\xi) = \int_{q_*}^{q^*} \left(-\frac{R}{q(1-R)} \frac{O(q,n(q))}{n(q)} \right) dq \qquad (3.11)$$

holds.

In case where $1 \in [p_*, p^*]$ or equivalently $1 \in [q_*, q^*]$, the integrals $\int_{p_*}^{p^*} \frac{dp}{p(1-p)}$ and $\int_{q_*}^{q^*} \frac{dq}{q(1-q)}$ are not well defined. But it can be shown that (3.11) still holds using a limiting argument, see Appendix 3.H.

To summarise, we would like to solve the following:

(The free boundary value problem) find a positive function $n(\cdot)$ and a pair of boundary points (q_*, q^*) solving

$$\begin{aligned} n'(q) &= O(q, n(q)), \qquad q \in [q_*, q^*] \\ n(q_*) &= m(q_*), \qquad n(q^*) = m(q^*) \end{aligned} \tag{3.12}$$

subject to (3.11).

In Section 3.5, we distinguish several different cases and discuss how to construct the solution $(n(\cdot), q_*, q^*)$ in each of these cases. The importance of the quadratic m is clear from (3.12). Meanwhile, the function ℓ acts as a bound on the feasible solution to $n' = O(q, n)$, at least on $0 < q \leqslant 1$. Suppose, for example, that $R < 1$. Then for $q \in [q_*, q^*]$, $m(q) \leqslant n(q)$ by construction, and we also have $n(q) < \ell(q)$ on $q_* \leqslant q \leqslant 1$. Furthermore, the value of $\ell(1)$ is key in determining when the problem is ill-posed.

3.4 Main results

In Section 3.3 we converted the original HJB equation into the free boundary value problem (3.12). Now we argue that, given a solution $(n(\cdot), q_*, q^*)$ to (3.12) we can reverse the transformations and construct a candidate value function.

Suppose there exists a solution $(n(\cdot), q_*, q^*)$ to (3.12) with n being strictly positive. Define $p_* = \frac{q_*}{1+\lambda(1-q_*)}$ and $p^* = \frac{q^*}{1-\gamma(1-q^*)}$. Let $N(q) = \text{sgn}(1-q)n(q)^{-R}|1-q|^{R-1}$, $W = N^{-1}$ and $w(h) = (1-R)hW(h)$. We would like to construct the candidate value function from $G(p) = \text{sgn}(1-p)|1-p|^{1-R}h(p)$ where h solves $\frac{dh}{dp} = \frac{w(h)}{p(1-p)}$. The main subtlety is that $\frac{w(h)}{p(1-p)}$ is not well defined at $p = 1$. Nonetheless, the definition of G at $p = 1$ can be understood in a limiting sense. To this end, we distinguish two different cases based on whether $(q_* - 1)$ and $(q^* - 1)$ have the same sign or not, or equivalently whether the no-transaction wedge, plotted in $(x, y\theta)$ space, includes the vertical axis $x = 0$ (corresponding to $p = 1$).

Proposition 3.3. *(i) For $1 \notin [p_*, p^*]$, define $h(p)$ via*

$$\int_{N(q_*)}^{h(p)} \frac{du}{w(u)} = \int_{p_*}^{p} \frac{du}{u(1-u)} \tag{3.13}$$

on $p_* \leqslant p \leqslant p^*$. *Then* (3.13) *is equivalent to*

$$\int_{h(p)}^{N(q^*)} \frac{du}{w(u)} = \int_{p}^{p^*} \frac{du}{u(1-u)} \tag{3.14}$$

and (3.14) *is an alternative definition of* $h(p)$.

Let

$$G^C(p) = \begin{cases} n(q_*)^{-R} (1+\lambda p)^{1-R}, & p \in [-\frac{1}{\lambda}, p_*), \\ \operatorname{sgn}(1-p)|1-p|^{1-R} h(p), & p \in [p_*, p^*], \\ n(q^*)^{-R} (1-\gamma p)^{1-R}, & p \in (p^*, \frac{1}{\gamma}]. \end{cases}$$

Then G^C *is a* C^2 *function on* $(-\frac{1}{\lambda}, \frac{1}{\gamma})$. *Moreover* $\frac{(x+y\theta)^{1-R}}{1-R} G^C(\frac{y\theta}{x+y\theta})$ *is strictly increasing and strictly concave in* x.

(ii) For $1 \in [p_*, p^*]$, *define* $h(p)$ *via*

$$\begin{cases} \int_{N(q_*)}^{h(p)} \frac{du}{w(u)} = \int_{p_*}^{p} \frac{du}{u(1-u)}, & p_* \leqslant p < 1, \\ \int_{h(p)}^{N(q^*)} \frac{du}{w(u)} = \int_{p}^{p^*} \frac{du}{u(1-u)}, & 1 < p \leqslant p^*. \end{cases}$$

Let

$$G^C(p) = \begin{cases} n(q_*)^{-R} (1+\lambda p)^{1-R}, & p \in [-\frac{1}{\lambda}, p_*), \\ \operatorname{sgn}(1-p)|1-p|^{1-R} h(p), & p \in [p_*, p^*] \setminus \{1\}, \\ n(1)^{-R} e^{-(1-R)a}, & p = 1, \\ n(q^*)^{-R} (1-\gamma p)^{1-R}, & p \in (p^*, \frac{1}{\gamma}], \end{cases}$$

with $a := -\int_{q_*}^{1} \left(\frac{R}{q(1-R)} \frac{O(q,n(q))}{n(q)} \right) dq - \ln(1+\lambda)$. *Then* $|a| \leqslant \ln(1+\xi)$, *and* G^C *is a* C^2 *function on* $(-\frac{1}{\lambda}, \frac{1}{\gamma})$. *Moreover* $\frac{(x+y\theta)^{1-R}}{1-R} G^C(\frac{y\theta}{x+y\theta})$ *is strictly increasing and strictly concave in* x.

The first pair of main results of this chapter are summarised in the following two theorems. For a given set of risk aversion parameter R, discount factor δ and market parameters r, μ, σ, α, η, ρ, we say the problem is well-posed if the value function is finite on the interior of the solvency region for all values of the transaction costs λ and γ. We say the problem is ill-posed if the value function is infinite for all $\lambda > 0$ and $\gamma \in (0,1)$. We say the problem is conditionally well-posed if the value function is well-posed for large values of the round-trip transaction cost, but ill-posed for small values.

Theorem 3.4. *The investment/consumption problem is:*

1. *well-posed in either of the following cases:*

(a) $R > 1$;

(b) $R < 1$ and $m_M \geq 0$;

2. *ill-posed if $R < 1$, $m_M < 0$ and $\ell(1) \leqslant 0$;*

3. *conditionally well-posed if $R < 1$, $m_M < 0$ and $\ell(1) > 0$. In this case the problem is well-posed if and only if $\xi > \overline{\xi}$ where $\overline{\xi}$ is defined in* (3.18) *below.*

Note that, if $R < 1$ and $m_M = 0$ and $\lambda = 0 = \gamma$ (a case we have excluded) then the problem is ill-posed.

Theorem 3.5. *Suppose the parameters are such that the problem is well-posed. Set*

$$V^C(x,y,\theta) = \left(\frac{b_1}{Rb_4}\right)^{-R} \frac{(x+y\theta)^{1-R}}{1-R} G^C\left(\frac{y\theta}{x+y\theta}\right)$$

where G^C is as defined as in the relevant case of Proposition 3.3. Then $V^C = V$ where V is the value function of the investment/consumption problem defined in (3.1).

3.5 Solutions to the free boundary value problem

In this section we study the solutions to the free boundary problem (3.12) under several different combinations of parameters.

Let $\mathcal{S} \subseteq \{(q,n) : q > 0, n \geqslant 0\} = (0,\infty) \times [0,\infty)$ be the set $\mathcal{S} = \{q = 1\} \cup \{q = \frac{R}{R-1}\} \cup \{n = 0\} \cup \{q < 1, (1-R)n \geq (1-R)\ell(q)\}$. On $(0,\infty) \times [0,\infty) \setminus \mathcal{S}$ define $F(q,n) = O(q,n)/n$. Extend the definition of F to $(0,\infty) \times [0,\infty)$ where possible by taking appropriate limits. We begin this section with a list of useful results regarding the functions m and ℓ and operators O and F.

Lemma 3.6. 1. (a) *For $R < 1$, $\ell(q) > m(q)$ on $q \in (0,1]$. Moreover, on $(0,\infty)$, m crosses ℓ exactly once from below at some point above 1;*

(b) *For $R > 1$, $m(q) > \ell(q)$ on $q \in (0,1]$. Moreover, on $(0,\infty)$, m either does not cross ℓ at all, or touches ℓ exactly once in the open interval $(1, R/(R-1))$, or crosses ℓ twice on $(1, R/(R-1))$. Also, $\ell(q) > m(q)$ on $(R/(R-1), \infty)$;*

2. *For $R > 1$, $F(q,n)$ is well defined at $q = R/(R-1)$;*

3. *For $n > 0$ and $(1-R)n < (1-R)\ell(1)$, $F(1,n)$ is well-defined and*

$$F(1,n) := \lim_{q\to 1} F(q,n) = -\frac{(1-R)(n-m(1))}{\ell(1)-n}. \tag{3.15}$$

Also, for $0 < q \le 1$ *and* $R < 1$ *we have* $\lim_{n\uparrow \ell(q)} F(q,n) = -\infty$ *(and* $\lim_{n\downarrow \ell(q)} F(q,n) = +\infty$ *if* $R > 1$*). For* $q > 1$ *and* $R < 1$ *(and* $1 < q < \frac{R}{R-1}$ *for* $R > 1$*) we have*

$$F(q,\ell(q)) := \lim_{n\to \ell(q)} F(q,n) = -\frac{1-R}{R(1-q)} \left\{ \frac{q[(1-R)q+R]}{[(1-R)q+R]^2 + (b_2-1)R^2} - 1 \right\}. \tag{3.16}$$

4. $F(q,n) = 0$ *if and only if* $n = m(q)$*. Moreover,*

   (a) *for* $R < 1$*:*

      i. *On* $0 < q < 1$*,* $F(q,n) < 0$ *for* $m(q) < n < \ell(q)$ *and* $F(q,n) > 0$ *for* $n < m(q)$ *or* $n > \ell(q)$*;*

      ii. *At* $q = 1$*,* $F(1,n) < 0$ *for* $m(1) < n < \ell(1)$ *and* $F(1,n) > 0$ *for* $n < m(1)$*.* $F(1,n)$ *is not well-defined for* $n \geqslant \ell(1)$*;*

      iii. *On* $q > 1$*,* $F(q,n) < 0$ *for* $n > m(q)$ *and* $F(q,n) > 0$ *for* $n < m(q)$*;*

   (b) *for* $R > 1$*:*

      i. *On* $0 < q < 1$*,* $F(q,n) > 0$ *for* $\ell(q) < n < m(q)$ *and* $F(q,n) < 0$ *for* $n < \ell(q)$ *or* $n > m(q)$*;*

      ii. *At* $q = 1$*,* $F(1,n) > 0$ *for* $\ell(1) < n < m(1)$ *and* $F(1,n) < 0$ *for* $n > m(1)$*.* $F(1,n)$ *is not well-defined for* $n \leqslant \ell(1)$*;*

      iii. *On* $1 < q \leqslant R/(R-1)$*,* $F(q,n) < 0$ *for* $n > m(q)$ *and* $F(q,n) > 0$ *for* $n < m(q)$*;*

      iv. *On* $q > R/(R-1)$*,* $F(q,n) < 0$ *for* $m(q) < n < \ell(q)$ *and* $F(q,n) > 0$ *for* $n > \ell(q)$ *or* $n < m(q)$*.*

Recall (q_M, m_M) is the extrema of the quadratic m (a minima when $R < 1$ and a maxima when $R > 1$) with $q_M = \frac{b_3}{2R} > 0$. The key analytical properties of the problem only depend on the signs of the three parameters $(1-R, m_M, \ell(1))$. We classify four different cases using the decision tree in Figure 3.1.

We parameterise the family of solutions to (3.12) by the left boundary point. Fix $u \in (0, q_M)$ and denote $(n_u(q))_{q \geqslant u}$ the solution to the initial value problem

$$n'(q) = O(q, n(q)), \quad n(u) = m(u).$$

Note that $O(q,n)$ and $\frac{\partial O}{\partial n}(q,n)$ are well defined away from the set $\mathcal{S}$ (with appropriate extension by taking limits whenever applicable). Then standard theories assure the existence

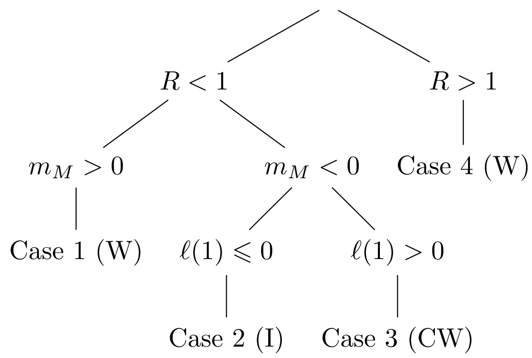


Figure 3.1: Classification of different cases based on the signs of the parameters. The abbreviations in parentheses indicate the solution features of the cases, where "W" refers to unconditional well-posedness for all levels of transaction cost, "I" refers to unconditional ill-posedness for all levels of transaction cost and "CW" refers to conditional well-posedness, i.e. well-posedness for sufficiently high levels of transaction cost only.

and uniqueness of the solution to the above initial value problem. Let $\zeta(u) = \inf\{q \geqslant u : (1-R)n_u(q) < (1-R)m(q)\}$ denote where n_u first crosses m to the right of u. Define

$$\Sigma(u) = \exp\left(\int_u^{\zeta(u)} \left(-\frac{R}{q(1-R)}\frac{O(q,n_u(q))}{n_u(q)}\right) dq\right) - 1. \tag{3.17}$$

Lemma 3.7. *Suppose $m_M > 0$. Then Σ is a strictly decreasing, continuous mapping $\Sigma : (0, q_M] \to [0,\infty)$ with $\Sigma(0+) = +\infty$ and $\Sigma(q_M) = 0$.*

Now suppose $m_M \leq 0$. Let $p_- \leq p_+$ be the roots of $m(q) = 0$. Set $\overline{\xi} := \lim_{u\uparrow p_-} \Sigma(u)$. Then Σ is a strictly decreasing, continuous mapping $\Sigma : (0, p_-] \to [\overline{\xi},\infty)$ with $\Sigma(0+) = +\infty$ and $\Sigma(p_-) = \overline{\xi}$. Moreover, $\lim_{u\uparrow p_-} n_u(\cdot) = 0$, $\lim_{u\uparrow p_-} \zeta(u) = p_+$, and

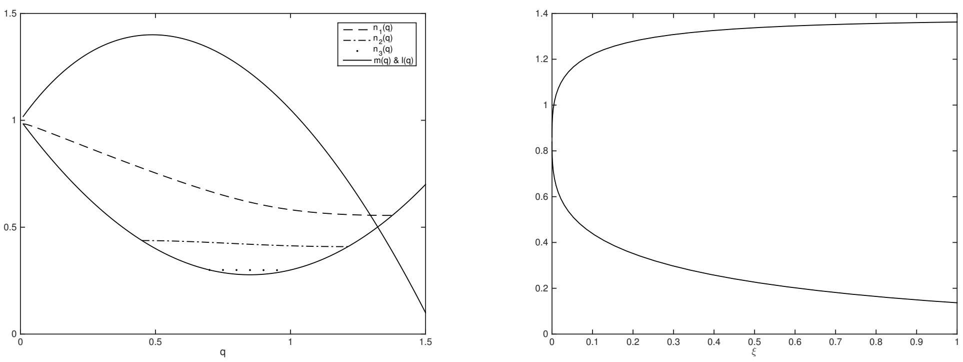
$$\overline{\xi} = \exp\left(-\int_{p_-}^{p_+} \frac{R}{q(1-R)} F(q,0) dq\right) - 1. \tag{3.18}$$

3.5.1 Case 1: $R < 1$ and $m_M \geq 0$

For any initial value $u \in (0, q_M)$, $m'(u) < 0 = O(u, m(u)) = O(u, n_u(u)) = n_u'(u)$. Thus $n_u(q)$ must initially be larger than $m(q)$ for q being close to u. By part 4 of Lemma 3.6, $O(q,n)$ is negative on $\{(q,n) : 0 < q \leqslant 1, m(q) < n < \ell(q)\} \cup \{(q,n) : q > 1, n > m(q)\}$. Also, $n_u(q)$ cannot cross $\ell(q)$ from below on $0 < q \leqslant 1$ since $\lim_{n\uparrow \ell(q)} O(q,n) = -\infty$. By considering the sign of $O(q,n)$, we conclude n_u must be decreasing until it crosses m. This guarantees the finiteness of $\zeta(u)$, and the triple $(n_u(\cdot), u, \zeta(u))$ represents one possible solution to problem (3.12). Notice that the family of solutions $(n_u(\cdot))_{0<u<q_M}$ cannot cross, and thus $n_u(q)$ is decreasing in u. The solutions corresponding to initial values $u = 0$ and $u = q_M$ can be understood as the appropriate limit of a sequence of solutions.

Although $O(q, n)$ has singularities at $q = 1$ and $n = \ell(q)$, part 3 of Lemma 3.6 shows that a well-defined limit $O(q, n)$ exists on $\{(q, n) : q = 1, n < \ell(1)\}$ and $\{(q, n) : q > 1, n = \ell(q)\}$. Hence there exists a continuous modification of $O(q, n)$ and a solution n_u can actually pass through these singularity curves. See Figure 3.2(a) for some examples.

From the analysis leading to (3.11), the correct choice of u should satisfy $\xi = \Sigma(u)$. From Lemma 3.7, for every given level of round-trip transaction cost ξ, there exists a unique choice of the left boundary point given by $u_* = \Sigma^{-1}(\xi)$ and then the desired solution to the free boundary value problem is given by $(n_{u_*}(\cdot), u_*, \zeta(u_*))$. Figure 3.2(b) gives the plots of $\Sigma^{-1}(\xi)$ and $\zeta(\Sigma^{-1}(\xi))$ representing the boundaries (q_*, q^*) under different levels of transaction costs.



(a) Examples of solutions $n_u(q)$ with different initial values $(u, m(u))$.

(b) Plots of $q_* = \Sigma^{-1}(\xi)$ and $q^* = \zeta(q_*)$.

Figure 3.2: Case 1 where parameters chosen are $R = 0.5$, $b_1 = 0.25$, $b_2 = 1.75$ and $b_3 = 0.85$.

3.5.2 Case 2: $R < 1$, $m_M < 0$, $\ell(1) \leq 0$

Let ℓ_0 be the root of $\ell(q) = 0$ on $q \in (0, 1)$. Since the solution to $n'(q) = O(q, n(q))$ must be bounded below by zero[1] and above by $\ell(q)$ for $q \in (0, \ell_0)$, for any initial value $(u, m(u))$

[1]Suppose a solution $\overline{n}$ starting at $(u, m(u))$ hits zero at some $q = q_0$ with $\ell(q_0) > 0$. Then both $\overline{n}$ and $n_0 = n_0(q) := 0$ are solutions to the initial value problem $n'(q) = O(q, n(q))$ with $n(q_0) = 0$, but this contradicts the uniqueness of the solution given that $\frac{\partial O}{\partial n}(q, n)$ is well defined along the trajectories of $\overline{n}(q)$ and $n_0(q)$ over $q \in [u, q_0]$.

for which $m(u) > 0$, the corresponding solution $n_u(\cdot)$ must hit $(\ell_0, 0)$. Hence there does not exist any strictly positive solution which crosses m again to the right of u. See Figure 3.3. In this case, there is no solution to the free boundary value problem and indeed the underlying problem is ill-posed for all levels of transaction costs and thus the value function cannot be defined.

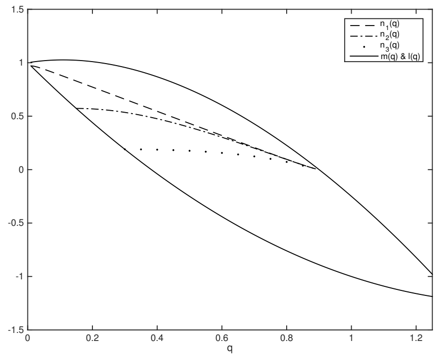


Figure 3.3: Case 2 where parameters chosen are $R = 0.5$, $b_1 = 0.25$, $b_2 = 1.75$ and $b_3 = 1.5$.

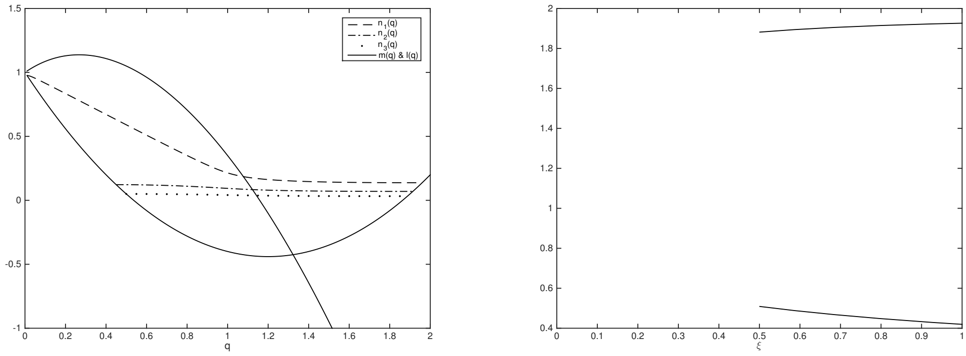
3.5.3 Case 3: $R < 1$, $m_M < 0$, $\ell(1) > 0$

Let $p_\pm$ with $0 < p_- < q_M < p_+$ be the two roots of $m(q) = 0$. The parameterisation of the solution is the same as in Case 1 except the left boundary point should now be restricted to $u \in (0, p_-)$ to ensure a positive initial value. The function Σ defined in (3.17) is still a strictly decreasing map with $\Sigma(0+) = +\infty$ except its domain is now restricted to $(0, p_-]$.

Unlike Case 1, we now only consider $\Sigma^{-1}(\xi)$ on the range $\xi \in (\bar{\xi}, \infty)$. For such a given high level of round-trip transaction cost, the required left boundary point is given by $u_* = \Sigma^{-1}(\xi)$ and $u^* = \zeta(u_*)$, see Figure 3.4. In this case, the problem is conditionally well-posed only for a sufficiently high level of transaction costs.

3.5.4 Case 4: $R > 1$.

In this case the quadratic m has a positive maxima at (q_M, m_M) and $m(q) > \ell(q)$ on $q \in (0, 1)$. By checking the sign of $O(q, n)$ using part 4 of Lemma 3.6, one can verify that the solution n_u of the initial value problem is always increasing for any choice of left boundary point $u \in (0, q_M)$. In this case the family of solutions is increasing in u. The solution $n_u(q)$ crosses $m(q)$ from below at $\zeta(u) = \inf(q \geqslant u : n_u(q) > m(q))$. The correct

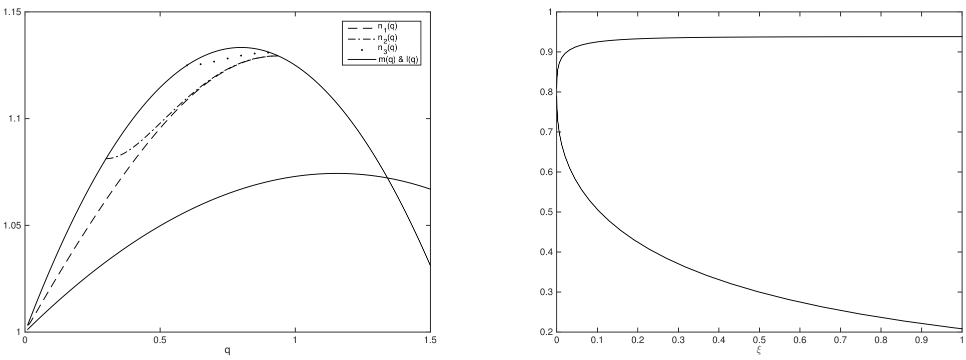


(a) Examples of solutions $n_u(q)$ with different initial values $(u, m(u))$.

(b) Plots of $q_* = \Sigma^{-1}(\xi)$ and $q^* = \zeta(q_*)$.

Figure 3.4: Case 3 where parameters chosen are $R = 0.5$, $b_1 = 0.25$, $b_2 = 1.75$ and $b_3 = 1.2$.

choice of u is again the one solving $\xi = \Sigma(u)$ using the same definition in (3.17). As in Case 1, the function Σ is onto from $(0, q_M]$ to $[0, \infty)$ and hence $u_* = \Sigma^{-1}(\xi)$ always exists uniquely for any ξ. See Figure 3.5. Indeed for $R > 1$, the agent's utility function is always bounded above by zero and hence the value function always exists and is finite.



(a) Examples of solutions $n_u(q)$ with different initial values $(u, m(u))$.

(b) Plots of $q_* = \Sigma^{-1}(\xi)$ and $q^* = \zeta(q_*)$.

Figure 3.5: Case 4 where parameters chosen are $R = 1.25$, $b_1 = 1.5$, $b_2 = 1.25$ and $b_3 = 2$.

3.6 Comparative statics

In this section, we investigate how the no-transaction wedge $[p_*, p^*]$, the value function V and the certainty equivalent value of the illiquid asset change with the market parameters and level of transaction costs.

3.6.1 Monotonicity with respect to market parameters

Proposition 3.8. *Suppose $(n(\cdot), q_*, q^*)$ is the solution to the free boundary value problem. Then:*

1. *q_* and q^* are decreasing in b_1;*
2. *For $R < 1$, q_* and q^* are increasing in b_3.*

Recall that $p_* = \frac{q_*}{1+\lambda(1-q_*)}$ and $p^* = \frac{q^*}{1-\gamma(1-q^*)}$. Then, Proposition 3.8 gives immediately:

Theorem 3.9. 1. *p_* and p^* are decreasing in b_1;*

2. *For $R < 1$, p_* and p^* are increasing in b_3.*

Theorem 3.9 describes the comparative statics in terms of the auxiliary parameters.[2] In general, it is difficult to make categorical statements about the comparative statics with respect to the original market parameters since many of the market parameters enter the definitions of more than one of the auxiliary parameters. However, we have the following results concerning the dependence of p_* and p^* on the discount rate, and on the drift of the illiquid asset.

Corollary 3.10. *p_* and p^* are decreasing in δ. If $R < 1$ then p_* and p^* are increasing in α.*

Now we consider the "cash value" of the holdings in the illiquid asset. We compare the agent with holdings in the illiquid asset to an otherwise identical agent (same risk aversion and discount parameter, and trading in the financial market with bond and risky asset with price S) who has a zero initial endowment in the illiquid asset and is precluded from taking any positions in this asset.

[2]Since the free boundary value problem does not depend on b_4, q_* and q^* are trivially independent of b_4. We have strong numerical evidence that p_* is decreasing in b_2 and p^* is increasing in b_2, and when $R > 1$ p_* and p^* are both increasing in b_3, but we have not been able to prove the results.

Consider the market without the illiquid asset. For an agent operating in this market a consumption/investment strategy is admissible for initial wealth $x > 0$ (we write $(C = (C_t)_{t\geq 0}, \Pi = (\Pi_t)_{t\geq 0}) \in \mathcal{A}_W(x)$) if C and Π are progressively measurable, and if the resulting wealth process $X = (X_t)_{t\geq 0}$ is non-negative for all t. Here X solves

$$dX_t = r(X_t - \Pi_t)dt + \frac{\Pi_t}{S_t}dS_t - C_t dt$$

subject to $X_0 = x$. Let $W = W(x)$ be the value function for a CRRA investor:

$$W(x) = \sup_{(C,\Pi)\in\mathcal{A}_W(x)} \mathbb{E}\left[\int_0^\infty e^{-\delta t}\frac{C_t^{1-R}}{1-R}dt\right].$$

The problem of finding W is a classical Merton consumption/investment problem without transaction costs. We find

$$W(x) = \left[\frac{1}{R}\left(\delta - r(1-R) - \frac{\beta^2(1-R)}{2R}\right)\right]^{-R}\frac{x^{1-R}}{1-R} = \left(\frac{b_1}{b_4 R}\right)^{-R}\frac{x^{1-R}}{1-R}.$$

Define $\mathcal{C} = \mathcal{C}(y\theta; x)$ to be the certainty equivalent value of the holding of the illiquid asset, i.e. the cash amount which the agent with liquid wealth x and θ units of the illiquid asset with current price y, trading in the market with transaction costs, would exchange for his holdings of the illiquid asset, if after this exchange he is not allowed to trade in the illiquid asset. (We assume there are no transaction costs on this exchange, but they can be easily added if required.) Then $\mathcal{C} = \mathcal{C}(y\theta; x)$ solves

$$W(x + \mathcal{C}) = V(x, y, \theta)$$

which becomes

$$\mathcal{C} = \mathcal{C}(y\theta; x) = (x + y\theta)G(p)^{1/(1-R)} - x.$$

Theorem 3.11. *1. $(1-R)G$ is decreasing in b_1;*

2. $(1-R)G$ is increasing in b_3.

Corollary 3.12. *$\mathcal{C}$ is decreasing in δ and increasing in α.*

Both these monotonicities are intuitively natural. For the monotonicity in α, since the agent only ever holds long[3] positions in the illiquid asset, we expect him to benefit from an increase in drift and hence price of the illiquid asset. If we consider monotonicity in δ then for $R < 1$, increasing δ reduces the magnitude of the discounted utility of consumption,

[3]Note, if he starts with a solvent initial portfolio, but with a negative holding in the illiquid asset, then the agent makes an instantaneous transaction at time zero to make his holding positive.

and reduces the value function. However, this is not the same as decreasing the certainty equivalent value of the holding of risky asset. Indeed, when $R > 1$, increasing δ reduces the magnitude of the discounted utility of consumption, but since the terms are negative, this increases the value function. Nonetheless, $\mathcal{C}$ is decreasing in δ.

3.6.2 Monotonicity with respect to transaction costs

From the discussion in Section 3.5, we have seen that transformed boundaries only depends on the round-trip transaction cost ξ. In particular, q_* and q^* are respectively strictly decreasing and increasing in ξ. However, the purchase/sale boundaries in the original scale still depend on the individual costs of purchase and sale. Write

$$p_*(\lambda,\gamma) = \frac{q_*(\xi)}{1+\lambda(1-q_*(\xi))}, \quad p^*(\lambda,\gamma) = \frac{q^*(\xi)}{1-\gamma(1-q^*(\xi))}$$

and recall that $\xi = \frac{\lambda+\gamma}{1-\gamma}$. Then

$$\frac{dp_*}{d\gamma} = \frac{\partial p_*}{\partial q_*}\frac{\partial q_*}{\partial \xi}\frac{\partial \xi}{\partial \gamma} = \frac{1+\lambda}{(1-\gamma)^2}\frac{1+\lambda}{[1+\lambda(1-q_*)]^2}\frac{\partial q_*}{\partial \xi} < 0$$

so that the critical ratio of wealth in the illiquid asset to paper wealth at which the agent purchases more illiquid asset is decreasing in the transaction cost on sales. However, perhaps surprisingly, the dependence of the critical ratio p_* at which purchases occur on the transaction cost on purchases is not unambiguous in sign:

$$\frac{dp_*}{d\lambda} = \frac{\partial p_*}{\partial \lambda} + \frac{\partial p_*}{\partial q_*}\frac{\partial q_*}{\partial \xi}\frac{\partial \xi}{\partial \lambda} = -\frac{q_*(1-q_*)}{[1+\lambda(1-q_*)]^2} + \frac{1}{1-\gamma}\frac{1+\lambda}{[1+\lambda(1-q_*)]^2}\frac{\partial q_*}{\partial \xi}$$

is not necessarily negative, for we may have $q_* > 1$. This issues is discussed further in Hobson et al. (2016) where examples are given in which the boundaries to the no-transaction region are not monotonic in the transaction cost parameters.

3.7 Conclusion

The solution of Merton (1969) of the infinite horizon, consumption and investment problem is elegant and insightful but assumes a perfect market with no frictions. Building on this work, there is a large literature, starting with Magill and Constantinides (1976) and Davis and Norman (1990) investigating the form of the solution in the presence of transaction costs. When there is a single asset Choi et al. (2013) (via shadow prices) and Hobson et al. (2016) (via an analysis of the HJB equation) are able to characterise precisely when the problem is well-posed. However, Davis and Norman (1990), Choi et al. (2013) and Hobson et al. (2016) all assume the financial market includes just a single risky asset.

In this chapter we have extended the results to two risky assets, and give a complete characterisation of the solution, but in the special case where transaction costs are payable on only one of the risky assets. The presence of the second risky asset, which may be used for hedging and investment purposes, makes the problem significantly more complicated than the single risky asset case, but we can extend the methods of Hobson et al. (2016) to give a complete solution. Indeed, up to evaluating an integral of a known algebraic function, we can determine exactly when the problem is well-posed and up to solving a free boundary value problem for a first order differential equation we can determine the boundaries of the no-transaction wedge.

At the heart of our analysis is this free boundary value problem. Although the utility maximisation problem depends on many parameters describing the agent (his risk aversion and discount rate), the market (the interest rate and the drifts, volatilities and correlations of the traded assets) and the frictions (the transaction costs on sales and purchases) the ODE depends on the risk aversion parameter and just three further parameters, and the solution we want can be specified further in terms of the round-trip transaction cost.

Building on the work of Choi et al. (2013), Hobson et al. (2016) give a solution to the problem in the case of a single risky asset. The major issue in Choi et al. (2013) and Hobson et al. (2016) is to understand the solution of an ODE as it passes through a singular point. In this chapter the problem is richer, and the ODE is more complicated, but in other ways the analysis is much simpler because although the key ODE has singularities, these can be removed.

In the chapter we have assumed a single illiquid asset and just one further risky asset, but the analysis extends immediately to the case of a single illiquid asset and several risky assets on which no transaction costs are payable, at the expense of a more complicated notation. This observation is a form of mutual fund theorem — the agent chooses to invest in the additional liquid financial assets in fixed proportions and these assets may be combined into a representative market asset. Details of the argument in a related context may be found in Evans et al. (2008). Nonetheless, the extension to a model with many risky assets with transaction costs payable on all of them remains a challenging open problem.

Appendix to Chapter 3

3.A Transformation of the HJB equation

Looking at the HJB Equation (3.5), and using intuition gained from similar problems, we expect that $V = V(x, y, \theta)$ can be written as $V(x, y, \theta) = x^{1-R} J(\frac{y\theta}{x})$ for J a function of a single variable z representing the ratio of wealth in the illiquid asset to wealth in the liquid assets. The equation for $J = J(z)$ contains expressions of the form $zJ'(z)$ and $z^2 J''(z)$ and so can be made into a homogeneous equation by the substitution $(z, J(z)) \mapsto (e^u, K(u))$. The second-order equation for K can then be reduced to a first order equation by setting $w(K) = \frac{dK}{du}$ and making K the subject of the equation, see Evans et al. (2008) or Hobson et al. (2016) for details of a similar order-reduction in a related problem. However, there are cases where $x = 0$ lies inside the no-transaction region and at this point z is undefined, and the above approach does not work. Hence, we need to use a different parametrisation. We use a parametrisation based on $P_t = \frac{Y_t \Theta_t}{X_t + Y_t \Theta_t}$ representing the proportion of paper wealth which is held in the illiquid asset. The delicate point at $x = \pm 0$ (or $z = \pm\infty$) becomes a delicate point at $p = 1$, but as we show by a careful analysis any singularities can be removed.

Using the form of value function in (3.2) to compute all the relevant partial derivatives, (3.5) can be rewritten as

$$
\begin{aligned}
0 = {} & \frac{b_1}{b_4}\left[G(p) - \frac{pG'(p)}{1-R}\right]^{1-1/R} - \delta G(p) + r(1-p)\left[(1-R)G(p) - pG'(p)\right] \\
& + \alpha\left[(1-R)pG(p) + p(1-p)G'(p)\right] \\
& + \frac{\eta^2}{2}\left[p^2(1-p)^2 G''(p) - 2Rp^2(1-p)G'(p) - R(1-R)p^2 G(p)\right] \\
& - \frac{\left\{\beta\left[(1-R)G(p) - pG'(p)\right] + \eta\rho\left[-R(1-R)pG(p) + Rp(2p-1)G'(p) - p^2(1-p)G''(p)\right]\right\}^2}{2\left[p^2 G''(p) + 2RpG'(p) - R(1-R)G(p)\right]}.
\end{aligned}
\tag{3.19}
$$

Let[4] $h(p) = \operatorname{sgn}(1-p)|1-p|^{R-1}G(p)$ and $w(h) = p(1-p)\frac{dh}{dp}$. Then

$$
\frac{w(h)}{p(1-p)} = \frac{dh}{dp} = \operatorname{sgn}(1-p)|1-p|^{R-1}\left[G'(p) + (1-R)\frac{G(p)}{1-p}\right] \tag{3.20}
$$

[4] The assumption $b_3 > 0$ means that the agent would like to hold positive quantities of the illiquid asset, and that the no-transaction wedge is contained in the half-space $p > 0$. To allow for $b_3 < 0$ it is necessary to consider $p < 0$. This case can be incorporated into the analysis by incorporating an extra factor of $\operatorname{sgn}(p)$ into the definition of h, so that $h(p) = \operatorname{sgn}(p(1-p))|1-p|^{R-1}G(p)$. This then leads to extra cases, but no new mathematics, and the problem can still be reduced to solving $n' = O(q, n)$ where O is given by (3.7), but now for $q < 0$.

and in turn

$$G'(p) = \frac{w(h)}{|p||1-p|^R} - (1-R)\frac{G(p)}{1-p}. \tag{3.21}$$

This gives

$$\begin{aligned} G(p) - \frac{pG'(p)}{1-R} &= G(p) - \frac{p}{1-R}\left[\frac{w(h)}{|p||1-p|^R} - (1-R)\frac{G(p)}{1-p}\right] \\ &= |1-p|^{-R}h\left(1 - \frac{w(h)}{(1-R)h}\right). \end{aligned} \tag{3.22}$$

We expect that $V_x > 0$ and hence that this expression is positive. It follows that $\operatorname{sgn}(1-p) = \operatorname{sgn}(h) = \operatorname{sgn}(1 - W(h))$. Then

$$\left(G(p) - \frac{pG'(p)}{1-R}\right)^{1-\frac{1}{R}} = \operatorname{sgn}(1-p)|1-p|^{1-R}h|h|^{-\frac{1}{R}}\left|1 - \frac{w(h)}{(1-R)h}\right|^{1-\frac{1}{R}}, \tag{3.23}$$

$$(1-p)\left[(1-R)G(p) - pG'(p)\right] = \operatorname{sgn}(1-p)|1-p|^{1-R}\left[(1-R)h - w(h)\right],$$

$$(1-R)G(p) - pG'(p) = \frac{\operatorname{sgn}(1-p)|1-p|^{1-R}}{1-p}(1-R)h\left(1 - \frac{w(h)}{(1-R)h}\right)$$

and

$$(1-R)pG(p) + p(1-p)G'(p) \quad = \quad \operatorname{sgn}(1-p)|1-p|^{1-R}w(h).$$

Taking a further derivative

$$\begin{aligned} w(h)w'(h) &= p(1-p)\frac{dh}{dp}\frac{d}{dh}w(h) = p(1-p)\frac{d}{dp}w(h) \\ &= p(1-p)\frac{d}{dp}\left\{\operatorname{sgn}(1-p)|1-p|^{R-1}\left[p(1-p)G'(p) + (1-R)pG(p)\right]\right\} \\ &= \operatorname{sgn}(1-p)|1-p|^{R-1} \\ &\qquad \times \left[p^2(1-p)^2G''(p) + p(1-p)(1-2Rp)G'(p) + (1-R)p(1-Rp)G(p)\right] \end{aligned}$$

and hence the second order terms in (3.19) can be rewritten as:

$$\begin{aligned} &p^2(1-p)^2G''(p) - 2Rp^2(1-p)G'(p) - R(1-R)p^2G(p) \\ &\qquad = \operatorname{sgn}(1-p)|1-p|^{1-R}w(h)(w'(h) - 1), \\ &-R(1-R)pG(p) + Rp(2p-1)G'(p) - p^2(1-p)G''(p) \\ &\qquad = -\frac{\operatorname{sgn}(1-p)|1-p|^{1-R}}{1-p}\left(w'(h)w(h) - (1-R)w(h)\right), \end{aligned}$$

and

$$p^2G''(p) + 2RpG'(p) - R(1-R)G(p) = \frac{\operatorname{sgn}(1-p)}{|1-p|^{1+R}}\left[w(h)w'(h) + (2R-1)w(h) - R(1-R)h\right]. \tag{3.24}$$

Substituting back into (3.19), and dividing through by $\mathrm{sgn}(1-p)|1-p|^{1-R}$ we obtain

$$
\begin{aligned}
0 = {} & \frac{b_1}{b_4} h|h|^{-1/R} \left| 1 - \frac{w(h)}{(1-R)h} \right|^{1-1/R} - \delta h \\
& + r\left[(1-R)h - w(h)\right] + \alpha w(h) + \frac{\eta^2}{2} w(h)(w'(h) - 1) \\
& - \frac{\left\{\beta(1-R)h\left(1 - \frac{w(h)}{(1-R)h}\right) - \eta\rho\left[w'(h)w(h) - (1-R)w(h)\right]\right\}^2}{2\left[w(h)w'(h) + (2R-1)w(h) - R(1-R)h\right]}.
\end{aligned} \tag{3.25}
$$

Recall the definitions $W(h) = \frac{w(h)}{(1-R)h}$, $N = W^{-1}$ and $n(q) = |N(q)|^{-1/R}|1-q|^{1-1/R}$. Then $w(N(q)) = (1-R)N(q)W(N(q)) = (1-R)qN(q)$. Put $h = N(q)$ in (3.25) and divide by h. Then we have

$$
\begin{aligned}
0 = {} & \frac{b_1}{b_4} n(q) - \delta + r(1-R)(1-q) + \alpha(1-R)q + \frac{\eta^2}{2}(1-R)\left[qw'(N(q)) - q\right] \\
& - \frac{1-R}{2} \frac{\left\{\beta(1-q) - \eta\rho\left[qw'(N(q)) - (1-R)q\right]\right\}^2}{qw'(N(q)) + (2R-1)q - R}.
\end{aligned} \tag{3.26}
$$

Recall the definitions of the auxiliary constants $(b_i)_{i=1,2,3,4}$ given at the very start of Section 3.3.2. Rearranging (3.26) and multiplying by b_4

$$
\begin{aligned}
0 \quad = \quad & (1-R)q^2(w'(N(q)))^2 \\
& \quad + \left[b_1 n(q) - \left[b_1 + b_2 R(1-R)\right] + (b_3 + 2R - 2)(1-R)q\right] qw'(N(q)) \\
& \quad + \left[(2R-1)(b_3 - 1) + R^2(1 - b_2)\right](1-R)q^2 \\
& \quad + \left[(1-2R)b_1 + R(1-R)b_2 - R(1-R)b_3\right] q \\
& \quad + b_1 R + b_1\left[(2R-1)q - R\right] n(q) \\
=: \quad & A(qw'(N(q)))^2 + B(qw'(N(q))) + C.
\end{aligned} \tag{3.27}
$$

This can be viewed as a quadratic equation in $qw'(N(q))$. Note that the coefficients A, B, C depend on the market parameters only through the auxiliary parameters b_1, b_2, b_3.

We want the root corresponding to $V_{xx} < 0$. This is equivalent to

$$\frac{1}{1-R} p^2 G''(p) + \frac{2R}{1-R} pG'(p) - RG(p) < 0. \tag{3.28}$$

Using (3.24) and the fact that $\mathrm{sgn}(1-p) = \mathrm{sgn}(h)$, and multiplying (3.28) by $|1-p|^{R+1}/|h|$ we find we want the solution for which

$$\frac{1}{(1-R)} \frac{1}{h}\{w(h)w'(h) + (2R-1)w(h) - R(1-R)h\} = \{qw'(N(q)) + (2R-1)q - R\} < 0. \tag{3.29}$$

Consider (3.26) and write $u = qw'(N(q))$. Then for fixed q (3.26) is of the form $(1-R)a_1 u - a_2 = (1-R)\frac{(a_3 u + a_4)^2}{(u - a_5)}$ where $(a_i)_{1 \le i \le 5}$ are constants with $a_1 > a_3^2$ and $a_5 =$

$R-(2R-1)q$. It is easily seen that this equation has two solutions, one on each side of $u=a_5$, and that from (3.29) the one we want is the smaller root. Thus

$$qw'(N(q)) = \frac{-B-\operatorname{sgn}(A)\sqrt{B^2-4AC}}{2A}$$

where A, B, C are the constants in (3.27). Note that $\operatorname{sgn}(A)=\operatorname{sgn}(1-R)$. Then, we have

$$\begin{aligned}
\frac{n'(q)}{n(q)} &= \frac{1-R}{R(1-q)} - \frac{1}{R}\frac{N'(q)}{N(q)} \\
&= \frac{1-R}{R(1-q)} - \frac{1-R}{R}\frac{q}{qw'(N(q))-(1-R)q^2} \\
&= \frac{1-R}{R(1-q)} - \frac{1-R}{R}\frac{2Aq}{-B-\operatorname{sgn}(A)\sqrt{B^2-4AC}-2A(1-R)q^2}.
\end{aligned}$$

After some algebra, we arrive at

$$n'(q) = \frac{(1-R)n(q)}{R(1-q)} - \frac{2(1-R)^2qn(q)/R}{2(1-R)(1-q)\left[(1-R)q+R\right]-\varphi(q,n(q))-\operatorname{sgn}(1-R)\sqrt{\varphi(q,n(q))^2+E(q)^2}}.$$

3.B Continuity and smoothness of the candidate value function

Proof of Case (i) of Proposition 3.3. We have

$$\begin{aligned}
\int_{N(q_*)}^{N(q^*)}\frac{du}{w(u)} - \int_{p_*}^{p^*}\frac{du}{u(1-u)} &= \int_{q_*}^{q^*}\left(\frac{N'(u)}{(1-R)uN(u)} - \frac{1}{u(1-u)}\right)du \\
&\quad + \int_{q_*}^{q^*}\frac{du}{u(1-u)} - \int_{p_*}^{p^*}\frac{du}{u(1-u)} \\
&= \int_{q_*}^{q^*}\left(-\frac{R}{u(1-R)}\frac{O(u,n(u))}{n(u)}\right)du - \ln(1+\xi) \\
&= 0
\end{aligned}$$

using (3.11) and this establishes the equivalence of (3.13) and (3.14).

Suppose we have a solution $(n(\cdot),q_*,q^*)$ to (3.12) with n being strictly positive. Let $N(q)=\operatorname{sgn}(1-q)n(q)^{-R}|1-q|^{R-1}$, $W=N^{-1}$ and $w(h)=(1-R)hW(h)$. We set $G^C(p)=\operatorname{sgn}(1-p)|1-p|^{1-R}h(p)$ where h solves $\frac{dh}{dp}=\frac{w(h)}{p(1-p)}$. For notational convenience (and to allow us to write derivatives as superscripts) write G as shorthand for G^C.

First we check that G is C^2. Outside the no-transaction interval this is immediate from the definition, and follows on (p_*,p^*) from the fact that n and n' are continuous. This property is inherited by the pair (w,w') and then on integration by the trio (h,h',h'') and finally (G,G',G'').

It remains to check the continuity of G, G' and G'' at p_* and p^*. We prove the continuity at p_*; the proofs at p^* are similar. Using $\frac{1-q^*}{1-p^*} = \frac{1}{1+\lambda p^*}$ for the penultimate equivalence, we have

$$\begin{aligned}
G(p_*+) &= \operatorname{sgn}(1-p_*)|1-p_*|^{1-R}h(p_*) \\
&= \operatorname{sgn}(1-p_*)|1-p_*|^{1-R}\operatorname{sgn}(1-q_*)n(q_*)^{-R}|1-q_*|^{R-1} \\
&= n(q_*)^{-R}(1+\lambda p_*)^{1-R} = G(p_*-).
\end{aligned}$$

Then continuity of G' at p_* follows from (3.22) where

$$\begin{aligned}
G(p_*+) - \frac{p_* G'(p_*+)}{1-R} = |1-p_*|^{-R}h_*(1-W(h_*)) &= \frac{G(p_*+)}{1-p_*}(1-q_*) \\
&= \frac{G(p_*)}{1+\lambda p_*} = G(p_*-) - \frac{p_* G'(p_*-)}{1-R}.
\end{aligned}$$

Finally, from (3.24),

$$\begin{aligned}
& p_*^2 G''(p_*+) + 2Rp_*G'(p_*+) - R(1-R)G(p_*+) \\
&= \frac{G(p_*+)}{(1-p_*)^2 h_*}\left[w(h_*)w'(h_*) + (2R-1)w(h_*) - R(1-R)h_*\right] \\
&= \frac{G(p_*)}{(1-p_*)^2}\left[(1-R)W(h_*)w'(h_*) + (2R-1)(1-R)W(h_*) - R(1-R)\right] \\
&= \frac{G(p_*)}{(1-p_*)^2}\left[(1-R)^2 q_*\left(q_* + \frac{N(q_*)}{N'(q_*)}\right) + (2R-1)(1-R)q_* - R(1-R)\right] \\
&= \frac{(1-R)G(p_*)}{(1-p_*)^2}\left[(1-R)q_*\left(q_* + \frac{1-q_*}{1-R}\right) + (2R-1)q_* - R\right] \\
&= -R(1-R)G(p_*)\left(\frac{1-q_*}{1-p_*}\right)^2 \\
&= -R(1-R)\frac{G(p_*)}{(1+\lambda p_*)^2} \\
&= p_*^2 G''(p_*-) + 2Rp_*G'(p_*-) - R(1-R)G(p_*-),
\end{aligned}$$

where we have used the fact that

$$0 = \frac{n'(q_*)}{n(q_*)} = \frac{1-R}{R(1-q_*)} - \frac{1}{R}\frac{N'(q_*)}{N(q_*)}.$$

Then we conclude that G'' is continuous at $p = p_*$.

Now we argue that $\frac{(x+y\theta)^{1-R}}{1-R}G(\frac{y\theta}{x+y\theta})$ is strictly increasing and strictly concave in x. Outside $[p_*, p^*]$ this is immediate form the definition. On $[p_*, p^*]$ the increasing property will follow if $G(p) - \frac{pG'(p)}{1-R} > 0$. But this is trivial since

$$\begin{aligned}
G(p) - \frac{pG'(p)}{1-R} = |1-p|^{-R}h(1-W(h)) &= |1-p|^{-R}N(q)(1-q) \\
&= |1-p|^{-R}|1-q|^{R}n(q)^{-R} > 0.
\end{aligned}$$

Meanwhile, $\frac{(x+y\theta)^{1-R}}{1-R}G(\frac{y\theta}{x+y\theta})$ is concave on $[p_*, p^*]$ is equivalent to (3.28), or, by the analysis leading to (3.29) to $qw'(N(q))+(2R-1)q-R<0$. But this follows from our choice of root in (3.27).

$\square$

Proof of Case (ii) of Proposition 3.3. Note that the integrand of $\int_{q_*}^{q^*}\left(\frac{R}{q(1-R)}\frac{O(q,n(q))}{n(q)}\right)dq$ is everywhere negative and therefore $\int_{q_*}^{1}\left(-\frac{R}{q(1-R)}\frac{O(q,n(q))}{n(q)}\right)dq$ exists in $[0,\ln(1+\xi)]$. Hence $-\ln(1+\xi)\leqslant a\leqslant \ln(1+\xi)$.

For $p\neq 1$, the C^2 smoothness of $G=G^C$ follows as in the first case of Proposition 3.3. We will focus on the case of $p=1$.

Suppose first that $p_*<1<p^*$. Continuity of G and G' at $p=1$ can be established if we can show that both

$$\lim_{p\to 1}\frac{1}{G(p)}\left(G(p)-\frac{pG'(p)}{1-R}\right)^{1-1/R}=n(1) \tag{3.30}$$

and

$$\lim_{p\to 1}\frac{pG'(p)}{(1-R)G(p)}=1-e^{a}. \tag{3.31}$$

Substituting (3.31) into (3.30) we recover the given value of $G(1)$.

Using (3.23) and the equivalence of $p\to 1$ and $q\to 1$ we have

$$\frac{1}{G(p)}\left(G(p)-\frac{pG'(p)}{1-R}\right)^{1-\frac{1}{R}}=|h|^{-\frac{1}{R}}|1-W(h)|^{1-\frac{1}{R}}=|N(q)|^{-\frac{1}{R}}|1-q|^{1-\frac{1}{R}}=n(q)\to n(1)$$

and (3.30) holds.

For (3.31) we have,

$$\frac{1-W(h(p))}{1-p}=\frac{(1-R)h(p)-p(1-p)h'(p)}{(1-R)(1-p)h(p)}=1-\frac{pG'(p)}{(1-R)G(p)}.$$

Suppose $p<1$. Then using the definition of $h(p)$,

$$\begin{aligned}
0&=\int_{N(q_*)}^{h(p)}\frac{du}{w(u)}-\int_{p_*}^{p}\frac{du}{u(1-u)}\\
&=\int_{q_*}^{W(h(p))}\frac{N'(q)dq}{(1-R)qN(q)}-\int_{p_*}^{p}\frac{du}{u(1-u)}\\
&=\int_{q_*}^{W(h(p))}\left(\frac{N'(q)}{(1-R)qN(q)}-\frac{1}{q(1-q)}\right)dq+\int_{q_*}^{W(h(p))}\frac{dq}{q(1-q)}-\int_{p_*}^{p}\frac{du}{u(1-u)}\\
&=\int_{q_*}^{W(h(p))}\left(-\frac{R}{u(1-R)}\frac{O(u,n(u))}{n(u)}\right)du-\int_{p_*}^{q_*}\frac{du}{u(1-u)}-\int_{W(h(p))}^{p}\frac{dq}{q(1-q)}\\
&=\int_{q_*}^{W(h(p))}\left(-\frac{R}{u(1-R)}\frac{O(u,n(u))}{n(u)}\right)du-\ln(1+\lambda)-\ln\left(\frac{p}{W(h(p))}\frac{1-W(h(p))}{1-p}\right).
\end{aligned}$$

Letting $p \uparrow 1$ and using the fact that $\lim_{p\to 1} W(h(p)) = 1$, we obtain

$$\lim_{p\uparrow 1} \frac{1 - W(h(p))}{1-p} = e^a. \tag{3.32}$$

A similar calculation for $p > 1$ gives $\lim_{p\downarrow 1} \frac{W(h(p))-1}{p-1} = e^a$ as well. Hence (3.31) holds. As a byproduct, we can establish

$$\lim_{p\to 1} G'(p) = (1-R)(1-e^a)G(1) = (1-R)(1-e^a)n(1)^{-R}e^{-(1-R)a}.$$

Consider now continuity of G'' at $p = 1$. We show $\lim_{p\to 1} G''(p)$ exists. Consider:

$$\begin{aligned}\frac{[(1-R)G(p) - pG'(p)]^2}{G(p)[p^2G''(p) + 2RpG'(p) - R(1-R)G(p)]} &= \frac{(1-R)^2h(1-W(h))^2}{w(h)w'(h) + (2R-1)w(h) - R(1-R)h}\\ &= \frac{(1-R)(1-q)^2}{(1-R)qN(q)/N'(q) - (1-q)[R + (1-R)q]}\\ &= \frac{(1-R)\,[1-R-R(1-q)n'(q)/n(q)]}{R[R+(1-R)q]n'(q)/n(q) - R(1-R)}.\end{aligned}$$

Then,

$$\begin{aligned}\lim_{p\to 1} \frac{[(1-R)G(p) - pG'(p)]^2}{G(p)[p^2G''(p) + 2RpG'(p) - R(1-R)G(p)]} &= \lim_{q\to 1} \frac{(1-R)\,[1-R-R(1-q)n'(q)/n(q)]}{R\{[R+(1-R)q]n'(q)/n(q) - (1-R)\}}\\ &= \frac{(1-R)^2}{R[n'(1)/n(1) - (1-R)]}. \end{aligned}\tag{3.33}$$

Note that $n'(1)/n(1) - (1-R) \neq 0$ since $\operatorname{sgn}(n'(1)) = -\operatorname{sgn}(1-R)$. The limit is thus always well defined and can be used to obtain an expression for $\lim_{p\to 1} G''(p)$.

Since $\frac{(x+y\theta)^{1-R}}{1-R}G(\frac{y\theta}{x+y\theta})$ is strictly increasing, strictly concave in x for both $p < 1$ and $p > 1$, these properties extend to $p = 1$ under the second order smoothness of G. Note that the monotonicity and concavity at $p = 1$ must be strict due to (3.31) and (3.33).

Finally we consider the case where $p_* = 1$ or $p^* = 1$. Suppose we are in the former scenario. Then to show the continuity of G at $p_* = 1$ it is sufficient to show that

$$n(q_*)^{-R}\,(1+\lambda)^{1-R} = n(1)^{-R}e^{-(1-R)a}.$$

But $q_* = 1$ when $p_* = 1$ and thus $a = -\ln(1+\lambda)$. The above expression then holds immediately. Values of $G'(1)$ and $G''(1)$ can again be inferred from (3.31) and (3.33). A similar result follows in the case $p^* = 1$.

□

3.C The candidate value function and the HJB variational inequality

In this section we verify that the candidate value function given in Proposition 3.3 solves the HJB variational inequality

$$\min\left(-\sup_{c>0,\pi}\mathcal{L}^{c,\pi}V^C, -\mathcal{M}V^C, -\mathcal{N}V^C\right) = 0 \tag{3.34}$$

where $\mathcal{L}$, $\mathcal{M}$ and $\mathcal{N}$ are the operators

$$\begin{aligned}
\mathcal{L}^{c,\pi}f &:= \frac{c^{1-R}}{1-R} - cf_x + \frac{\sigma^2}{2}f_{xx}\pi^2 + ((\mu-r)f_x + \sigma\eta\rho f_{xy}y)\pi \\
&\quad + rf_x x + \alpha f_y y + \frac{\eta^2}{2}f_{yy}y^2 - \delta f, \\
\mathcal{M}f &:= f_\theta - (1+\lambda)yf_x, \\
\mathcal{N}f &:= (1-\gamma)yf_x - f_\theta.
\end{aligned}$$

Note that for $f = f(x,y,\theta)$ which is strictly increasing and concave in x we have

$$\mathcal{L}^*f := \sup_{c>0,\pi}\mathcal{L}^{c,\pi}f = \frac{R}{1-R}f_x^{1-1/R} + rxf_x + \alpha yf_y + \frac{\eta^2}{2}y^2f_{yy} - \frac{(\beta f_x + \eta\rho yf_{xy})^2}{2f_{xx}} - \delta f$$

and thus it is equivalent to show that $\min\left(-\mathcal{L}^*V^C, -\mathcal{M}V^C, -\mathcal{N}V^C\right) = 0$. From construction of V^C, it is trivial that $\mathcal{L}^*V^C = 0$, $\mathcal{M}V^C = 0$ and $\mathcal{N}V^C = 0$ on the no-transaction region, purchase-region and sale-region respectively. Hence it remains to show that

$$\begin{cases}
\mathcal{L}^*V^C \leqslant 0, \quad \mathcal{N}V^C \leqslant 0, \quad -1/\lambda \leqslant p < p_*, \\
\mathcal{M}V^C \leqslant 0, \quad \mathcal{N}V^C \leqslant 0, \quad p_* \leqslant p \leqslant p^*, \\
\mathcal{L}^*V^C \leqslant 0, \quad \mathcal{M}V^C \leqslant 0, \quad p^* < p \leqslant 1/\gamma.
\end{cases}$$

On the purchase region $p \in [-1/\lambda, p_*)$, direct substitution reveals that

$$\mathcal{N}V^C = -\left(\frac{b_1}{Rb_4}\right)^{-R} n(q_*)^{-R}(\lambda+\gamma)y(x+y\theta)^{-R}(1+\lambda p)^{-R} \leqslant 0,$$

and

$$\mathcal{L}^*V^C = \frac{R(x+y\theta)^{1-R}}{1-R}\left(\frac{b_1}{Rb_4}\right)^{1-R}(1+\lambda p)^{1-R}n(q_*)^{-R}\left(m(q_*) - m\left(\frac{(1+\lambda)p}{1+\lambda p}\right)\right) \leqslant 0$$

where we have used the facts that $n(q_*) = m(q_*)$, $\frac{(1+\lambda)p}{1+\lambda p} < \frac{(1+\lambda)p_*}{1+\lambda p_*} = q_*$ and the quadratic $m(q)$ is decreasing (respectively increasing) over $q < q_* < q_M$ when $R < 1$ (respectively $R > 1$). Similar calculations can be performed on the sale region $p \in (p^*, 1/\gamma]$ to show that $\mathcal{M}V^C \leqslant 0$ and $\mathcal{L}^*V^C \leqslant 0$.

Now we show that $\mathcal{M}V^C \leqslant 0$ on the no-transaction region $p \in [p_*, p^*]$. The inequality $\mathcal{N}V^C \leqslant 0$ can be proved in an identical fashion. Again writing G as shorthand for G^C, we have

$$\mathcal{M}V^C = V_\theta^C - (1+\lambda)yV_x^C = \frac{pV^C}{\theta}\left[(1+\lambda p)\frac{G'(p)}{G(p)} - \lambda(1-R)\right].$$

Since $\operatorname{sgn}(V^C) = \operatorname{sgn}(1-R)$, it is necessary and sufficient to show

$$\operatorname{sgn}(1-R)\left[(1+\lambda p)\frac{G'(p)}{G(p)} - \lambda(1-R)\right] \leqslant 0.$$

But $G(p) = \operatorname{sgn}(1-p)h(p)|1-p|^{1-R}$ for $p \neq 1$, and then

$$\frac{G'(p)}{G(p)} = \frac{h'(p)}{h(p)} - \frac{1-R}{1-p} = \frac{w(h)}{h(p)p(1-p)} - \frac{1-R}{1-p} = \frac{1-R}{1-p}\left(\frac{W(h)}{p} - 1\right)$$

and the required inequality becomes

$$\frac{1-W(h)}{1-p} \geqslant \frac{1}{1+\lambda p}. \tag{3.35}$$

We are going to prove (3.35) for $p \in [p_*, p^*] \setminus \{1\}$. Then $\mathcal{M}V^C \leqslant 0$ will hold at $p = 1$ as well by smoothness of V^C.

Suppose $p_* < p^* < 1$. Starting from the identity

$$\int_{N(q_*)}^{N(q)} \frac{dh}{w(h)} = \int_{p_*}^{p} \frac{du}{u(1-u)}$$

and following the substitutions leading to (3.10), we find

$$\int_{q_*}^{q}\left(-\frac{R}{u(1-R)}\frac{O(u,n(u))}{n(u)}\right)du = -\int_{q_*}^{q}\frac{dv}{v(1-v)} + \int_{p_*}^{p}\frac{du}{u(1-u)}.$$

Since the expression on the left hand side is increasing in q, we deduce

$$\frac{1}{p(1-p)}\frac{dp}{dq} \geqslant \frac{1}{q(1-q)}.$$

Define $\chi(q) := \frac{q}{1+\lambda(1-q)}$. Then χ is a solution to the ODE $\chi'(q) = \varrho(q, \chi(q))$ where $\varrho(q,y) = \frac{y(1-y)}{q(1-q)}$. Note that $\chi(q_*) = \frac{q_*}{1+\lambda(1-q_*)} = p_* = p(q_*)$. For $p_* \leqslant p \leqslant p^* < 1$ or equivalently $q_* \leqslant q \leqslant q^* < 1$, we have $p'(q) \geqslant \varrho(q, p(q))$, and we conclude $p(q) \geqslant \chi(q)$ for $q \in [q_*, q^*]$. Then

$$p = p(q) \geqslant \frac{q}{1+\lambda(1-q)} = \frac{W(h)}{1+\lambda(1-W(h))}$$

which establishes (3.35).

If instead $1 < p_* < p^*$, we can arrive at the same result by showing $p(q) \leqslant \chi(q)$ for $1 < q_* \leqslant q \leqslant q^*$ starting with $\int_{N(q)}^{N(q^*)} \frac{dh}{w(h)} = \int_{p}^{p^*} \frac{du}{u(1-u)}$ and then apply the same argument as in the previous case.

It remains to consider the case of $p_* \leqslant 1 \leqslant p^*$. The only issue is that the comparison of derivatives of $p(q)$ and $\chi(q)$ may not be trivial at $q = 1$ because of the singularity in $\varrho(q, y)$. But by direct computation, we find $\chi'(1) = 1 + \lambda$. On the other hand, consider an inverse expression $q = q(p)$ we have

$$q'(1-) = \lim_{p \uparrow 1} \frac{1 - q(p)}{1 - p} = \lim_{p \uparrow 1} \frac{1 - W(h(p))}{1 - p} = e^a$$

due to (3.32) and similarly we have $q'(1+) = e^a$. Then $q'(1)$ and in turn $p'(1) = 1/q'(1)$ is well-defined, and moreover since $a > -\ln(1 + \lambda)$ we have

$$p'(1) = e^{-a} < 1 + \lambda = \chi'(1).$$

Together with the fact that $p(1) = 1 = \chi(1)$, we must have that $\chi(q)$ is an upcrossing of $p(q)$ at $q = 1$. From this we conclude $p(q) \geqslant \chi(q)$ on $q \in [q_*, 1)$ and $\chi(q) \geqslant p(q)$ on $q \in (1, q^*]$. (3.35) then follows.

3.D Proof of the main results

Proof of Theorems 3.4 and 3.5. We prove the two theorems together. Suppose we are in the well-posed cases. From the analysis in Section 3.5, there exists a solution $(n(\cdot), q_*, q^*)$ to the free boundary value problem with n being strictly positive. By the C^2 smoothness of G^C, V^C is $C^{2\times 2\times 1}$. Moreover, in Appendices 3.B and 3.C we show that V^C is a strictly increasing and concave function in x solving the HJB variational inequality (3.34).

Let $M_t := \int_0^t e^{-\delta s} \frac{C_s^{1-R}}{1-R} ds + e^{-\delta t} V^C(X_t, Y_t, \Theta_t)$. Applying Ito's lemma, we obtain

$$\begin{aligned} M_t &= M_0 + \int_0^t e^{-\delta s} \mathcal{L}^{C_s, \Pi_s} V^C ds + \int_0^t e^{-\delta s} \mathcal{M} V^C d\Phi_s + \int_0^t e^{-\delta s} \mathcal{N} V^C d\Psi_s \\ &\quad + \int_0^t e^{-\delta s} \sigma V_x^C \Pi_s dB_s + \int_0^t e^{-\delta s} \eta V_y^C Y_s dW_s \\ &\leqslant M_0 + \int_0^t e^{-\delta s} \sigma V_x^C \Pi_s dB_s + \int_0^t e^{-\delta s} \eta V_y^C Y_s dW_s. \end{aligned}$$

Suppose $R < 1$. Then $M_t \geqslant 0$, and the sum of the stochastic integrals is a local martingale bounded below by $-M_0$ and in turn it is a supermartingale. Thus $\mathbb{E}(M_t) \leqslant M_0 = V^C(x, y, \theta)$ which gives

$$\mathbb{E}\left(\int_0^t e^{-\delta s} \frac{C_s^{1-R}}{1-R} ds \right) \leqslant V^C(x, y, \theta) - \mathbb{E}\left(e^{-\delta t} V^C(X_t, Y_t, \Theta_t) \right) \leqslant V^C(x, y, \theta).$$

On sending $t \to \infty$, we obtain $\mathbb{E}\left(\int_0^\infty e^{-\delta s} \frac{C_s^{1-R}}{1-R} ds \right) \leqslant V^C$ by monotone convergence and thus $V \leqslant V^C$ since C is arbitrary.

If $R > 1$, then the above argument does not go through directly since the local martingale will not be bounded below. But using the argument of Davis and Norman (1990), we can consider a perturbed candidate value function which is bounded on the no-transaction region and define a version of the value process M which will be a supermartingale. The result can be obtained by considering the limit of the perturbed candidate value function.

To show $V^C \leqslant V$, it is sufficient to demonstrate the existence of a strategy which attains the value V^C. Suppose the initial value $(x, y\theta)$ is such that $\frac{y\theta}{x+y\theta} = p \in [p_*, p^*]$. Define feedback controls $C^* = (C^*_t)_{t\geqslant 0}$ and $\Pi^* = (\Pi^*_t)_{t\geqslant 0}$ with $C^*_t = C^*(X_t, Y_t, \Theta_t)$ and $\Pi^*_t = \Pi^*(X_t, Y_t, \Theta_t)$ where

$$C^*(x,y,\theta) := [V^C_x(x,y,\theta)]^{-\frac{1}{R}}, \quad \Pi^*(x,y,\theta) := -\frac{(\mu - r)V^C_x(x,y,\theta_t) + \sigma\eta\rho y V^C_{xy}(x,y,\theta_t)}{\sigma^2 V^C_{xx}(x,y,\theta_t)},$$

and $\Theta^* = (\Theta^*_t)_{t\geqslant 0}$ a finite variation, local time strategy in form of $\Theta^*_t = \theta + \Phi^*_t - \Psi^*_t$ which keeps P_t within (p_*, p^*). Let X^* be the liquid wealth process evolving under these controls. Now since $(X^*, Y\Theta^*)$ is always confined in the no-transaction wedge, this strategy is clearly admissible.

Let M^* be the process $M = (M_t)_{t\geqslant 0}$ evolving under this controlled system. Then

$$\begin{aligned} M^*_t = M^*_0 &+ \int_0^t e^{-\delta s}\mathcal{L}^{C^*_s,\Pi^*_s}V^C ds + \int_0^t e^{-\delta s}\mathcal{M}V^C d\Phi^*_s + \int_0^t e^{-\delta s}\mathcal{N}V^C d\Psi^*_s \\ &+ \int_0^t e^{-\delta s}\sigma V^C_x \Pi^*_s dB_s + \int_0^t e^{-\delta s}\eta V^C_y Y_s dW_s \\ =:&\ M^*_0 + N^1_t + N^2_t + N^3_t + N^4_t + N^5_t. \end{aligned}$$

By construction of C^* and Π^*, $N^1_t = 0$. Moreover, Φ^* is carried by the set $\{P_t = p_*\}$ over which $\mathcal{M}V^C_s = 0$. Hence $N^2_t = 0$, and similarly $N^3_t = 0$. We show in Appendix 3.E that the local-martingale stochastic integrals N^4 and N^5 are martingales. Then on taking expectation we have

$$\mathbb{E}\left(\int_0^t e^{-\delta s}\frac{(C^*_s)^{1-R}}{1-R}ds\right) + \mathbb{E}(e^{-\delta t}V^C(X^*_t, Y_t, \Theta^*_t)) = \mathbb{E}(M^*_t) = M^*_0 = V^C. \tag{3.36}$$

We also show in Appendix 3.E that $\lim_{t\to\infty}\mathbb{E}(e^{-\delta t}V^C(X^*_t, Y_t, \Theta^*_t)) = 0$. Then letting $t \to \infty$ in (3.36) gives

$$V^C = \mathbb{E}\left(\int_0^\infty e^{-\delta s}\frac{(C^*_s)^{1-R}}{1-R}ds\right) \leqslant \sup_{(C,\Pi,\Theta)\in\mathcal{A}(0,x,y,\theta)} \mathbb{E}\left(\int_0^\infty e^{-\delta s}\frac{C_s^{1-R}}{1-R}ds\right) = V.$$

Now suppose the initial value $(x, y\theta)$ is such that $p < p_*$. Then consider a strategy of purchasing $\phi = \frac{xp_* - (1-p_*)y\theta}{y(1+\lambda p_*)}$ number of shares at time zero such that the post-transaction proportional holding in the illiquid asset is $\frac{y(\theta+\phi)}{x+y(\theta+\phi)-y(1+\lambda)\phi} = p_*$, and then follow the investment/consumption strategy (C^*, Π^*, Θ^*) as in the case of $p \in [p_*, p^*]$ thereafter. By construction of V^C, $V^C(x, y, \theta) = V^C(x - y(1+\lambda)\phi, y, \theta + \phi)$. Using (3.36) we have

$$\mathbb{E}\left(\int_0^t e^{-\delta s}\frac{(C_s^*)^{1-R}}{1-R}ds\right) + \mathbb{E}(e^{-\delta t}V^C(X_t^*, Y_t, \Theta_t^*)) = V^C(x - y(1+\lambda)\phi, y, \theta + \phi) = V^C(x, y, \theta)$$

and from this we can conclude $V^C \leqslant V$. Similar argument applies for initial value $p > p^*$.

Now we consider the set of parameters which leads to unconditional ill-posedness. It is sufficient to show that the problem without the liquid asset (which is the classical transaction cost problem involving one single risky asset only) is ill-posed. Note that $\ell(1) \leqslant 0$ is equivalent to $b_3 \geqslant \frac{b_1}{1-R} + b_2 R$ and this inequality can be restated as $\alpha \geqslant \frac{1}{2}\eta^2 R + \frac{\delta}{1-R}$. But this is exactly the ill-posedness condition in the one risky asset case. See Hobson et al. (2016) or Choi et al. (2013).

Finally we consider the conditionally well-posed case. From the discussion in Section 3.5, it is clear that as long as $\xi > \overline{\xi}$ there still exists $(n(\cdot), q_*, q^*)$ a solution to the free boundary value problem and thus one could show $V^C = V$ following the same argument in the proof for the unconditionally well-posed cases. Moreover, from Lemma 3.7 we can see that $n(\cdot) \downarrow 0$ as $\xi \downarrow \overline{\xi}$, in turn $V^C \to \infty$ from its construction. But $V \geqslant V^C$ and thus we conclude $V \to \infty$ as $\xi \downarrow \overline{\xi}$. This shows the ill-posedness of the problem at $\xi = \overline{\xi}$, and using the monotonicity of V in ξ this conclusion extends to any $\xi \leqslant \overline{\xi}$.

□

3.E The martingale property of the value process under the optimal control

Firstly we want to show that the $\frac{\sigma\Pi^* V_x^C}{V^C}$ and $\frac{\eta y V_y^C}{V^C}$ are bounded on the no-transaction region. Writing G^C as shorthand for G,

$$\begin{aligned}
\frac{\sigma\Pi^* V_x^C}{V^C} &= -\frac{V_x^C}{V^C}\frac{\beta V_x^C + \eta\rho y V_{xy}^C}{V_{xx}^C} \\
&= -\frac{(1-R)G(p) - pG'(p)}{G(p)} \\
&\quad \times \frac{\beta[(1-R)G(p) - pG'(p)] + \eta\rho[-R(1-R)pG(p) + Rp(2p-1)G'(p) - p^2(1-p)G''(p)]}{p^2G''(p) + 2RpG'(p) - R(1-R)G(p)} \\
&= -(1-R)\left(1 - \frac{w(h)}{(1-R)h}\right)\frac{\beta(1-R)h\left(1-\frac{w(h)}{(1-R)h}\right) - \eta\rho[w'(h)w(h) - (1-R)w(h)]}{w'(h)w(h) + (2R-1)w(h) - R(1-R)h} \\
&= -(1-R)(1-q)\frac{\beta(1-q) - \eta\rho\left[(1-R)q\left(q + \frac{N(q)}{N'(q)}\right) - (1-R)q\right]}{(1-R)q\frac{N(q)}{N'(q)} - (1-q)[R + (1-R)q]} \\
&= \frac{1-R}{R}(1-q)\left\{\beta - \frac{(1-R)(\eta\rho R - \beta)q\left(\frac{N(q)}{N'(q)} - (1-q)\right)}{R(1-q) - (1-R)q\left(\frac{N(q)}{N'(q)} - (1-q)\right)}\right\} \\
&= \frac{1-R}{R}(1-q)\left\{\beta - (\eta\rho R - \beta)\left[\frac{R(1-q)}{R(1-q) - (1-R)q\left(\frac{N(q)}{N'(q)} - (1-q)\right)} - 1\right]\right\}.
\end{aligned}$$

Using the monotonicity property of n it can be easily shown that $(1-R)(\frac{N(q)}{N'(q)} - \frac{1-q}{1-R}) \leq 0$, and then

$$R(1-q) - (1-R)q\left(\frac{N(q)}{N'(q)} - (1-q)\right) > R(1-q) - (1-R)q\left(\frac{1-q}{1-R} - (1-q)\right) = R(1-q)^2 \geqslant 0$$

and in turn $|R(1-q) - (1-R)q\left(\frac{N(q)}{N'(q)} - (1-q)\right)| > R(1-q)^2$. Thus

$$\begin{aligned}
\left|\frac{\sigma\Pi^* V_x^C}{V^C}\right| &\leqslant \frac{|1-R|}{R}|1-q|\left\{|\beta| + |\eta\rho R - \beta|\left[\frac{R|1-q|}{R(1-q) - (1-R)q\left(\frac{N(q)}{N'(q)} - (1-q)\right)} + 1\right]\right\} \\
&\leqslant \frac{|1-R|}{R}|1-q|\left\{|\beta| + |\eta\rho R - \beta|\left[\frac{R|1-q|}{R(1-q)^2} + 1\right]\right\} \\
&= \frac{|1-R|}{R}\left\{|\beta||1-q| + |\eta\rho R - \beta|\left[1 + |1-q|\right]\right\} \\
&\leqslant K_1
\end{aligned}$$

for some $K_1 > 0$ since $|1-q|$ is bounded on $q \in [q_*, q^*]$.

On the other hand,

$$\frac{\eta y V_y^C}{V^C} = \frac{\eta}{G(p)}[(1-R)pG(p) + p(1-p)G'(p)] = \eta\frac{w(h)}{h} = \eta(1-R)W(h)$$

and then $|\frac{\eta y V_y^C}{V^C}| \leqslant \eta|1-R|q^* =: K_2$.

Recall the notation $\mathcal{V}^C(x,y,\theta,t) = e^{-\delta t}V^C(x,y,\theta)$, and define a process $D = (D_t)_{t\geqslant 0}$ via $D_t = \ln \mathcal{V}^C(X_t^*, Y_t, \Theta_t^*, t) = \ln V^C(X_t^*, Y_t, \Theta_t^*) - \delta t$. Then

$$
\begin{aligned}
D_t = D_0 &- \int_0^t \frac{1}{1-R}\frac{(V_x^C)^{1-1/R}}{V^C}ds + \int_0^t \frac{\sigma \Pi_s^* V_x^C}{V^C}dB_s + \int_0^t \frac{\eta Y_s V_y^C}{V^C}dW_s \\
&- \frac{1}{2}\int_0^t \frac{1}{(V^C)^2}\left[\sigma^2(\Pi_s^*)^2(V_x^C)^2 + 2\sigma\eta\rho Y_s \Pi_s^* V_x^C V_y^C + \eta^2 Y_s^2 (V_y^C)^2\right] ds
\end{aligned}
$$

and hence the value process admits a representation of

$$
\mathcal{V}^C(X_t^*, Y_t, \Theta_t^*, t) = V^C(x,y,\theta)\exp\left(-\int_0^t \frac{1}{1-R}\frac{(V_x^C)^{1-1/R}}{V^C}ds\right) H_t
$$

with

$$
H_t := \mathcal{E}\left(\int_0^t \frac{\sigma \Pi_s^* V_x}{V^C}dB_s + \int_0^t \frac{\eta Y_s V_y^C}{V^C}dW_s\right)
$$

where $\mathcal{E}(Z_t) := \exp\left(Z_t - \frac{1}{2}[Z]_t\right)$ is the Doleans exponential of a process Z. Since $\int_0^t \frac{\sigma\Pi_s^* V_x}{V^C}dB_s$ and $\int_0^t \frac{\eta Y_s V_y^C}{V^C}dW_s$ have bounded integrands, H is a true martingale. Now,

$$
\begin{aligned}
(e^{-\delta t}\sigma\Pi_t^* V_x^C)^2 &\leqslant K_1^2 V^C(X_t, Y_t, \Theta_t)^2 \\
&\leqslant K_1^2 V^C(x,y,\theta)^2 H_t^2 \\
&= K_1^2 V^C(x,y,\theta)^2 \mathcal{E}\left(2\int_0^t \frac{\sigma\Pi_s^* V_x}{V^C}dB_s + 2\int_0^t \frac{\eta Y_s V_y^C}{V^C}dW_s\right) \\
&\quad\times \exp\left(\int_0^t \frac{1}{(V^C)^2}\left[\sigma^2(\Pi_s^*)^2(V_x^C)^2 + 2\sigma\eta\rho Y_s\Pi_s^* V_x^C V_y^C + \eta^2 Y_s^2(V_y^C)^2\right]ds\right) \\
&\leqslant K_1^2 V^C(x,y,\theta)^2 \mathcal{E}\left(2\int_0^t \frac{\sigma\Pi_s^* V_x}{V^C}dB_s + 2\int_0^t \frac{\eta Y_s V_y^C}{V^C}dW_s\right) \\
&\quad\times \exp\left(\left[K_1^2 + 2\rho K_1 K_2 + K_2^2\right]t\right)
\end{aligned}
$$

and therefore

$$
\mathbb{E}\left((e^{-\delta t}\sigma\Pi_t^* V_x^C)^2\right) \leqslant K_1^2 V^C(x,y,\theta)^2 \exp\left(\left[K_1^2 + 2\rho K_1 K_2 + K_2^2\right]t\right).
$$

We conclude $\mathbb{E}\left(\int_0^t (e^{-\delta s}\sigma\Pi_s^* V_x^C)^2 ds\right) < \infty$ and hence $N_t^4 = \int_0^t e^{-\delta s}\sigma\Pi_s^* V_x^C dB_s$ is a true martingale. Similarly, we can show that $N_t^5 = \int_0^t e^{-\delta s}\eta V_y^C Y_s dW_s$ is also a true martingale using the fact that $(e^{-\delta t}\eta V_y^C Y_t)^2 \leqslant K_2^2 V^C(x,y,\theta)^2 H_t^2$.

Finally, we want to show that $e^{-\delta t}V^C(X_t^*, Y_t, \Theta_t^*)$ converges to zero in L^1. We have

$$
\frac{1}{1-R}\frac{(V_x^C)^{1-1/R}}{V^C} = \frac{1}{G(p)}\left(G(p) - \frac{pG'(p)}{1-R}\right)^{1-1/R} = n(q) \geqslant n(q^*).
$$

Then

$$e^{-\delta t}V^C(X_t^*, Y_t, \Theta_t^*) = V^C(x,y,\theta)\exp\left(-\int_0^t \frac{1}{1-R}\frac{(V_x^C)^{1-1/R}}{V^C}ds\right)H_t$$
$$\leqslant V^C(x,y,\theta)\exp\left(-n(q^*)t\right)H_t$$

and hence $\lim_{t\to\infty}\mathbb{E}(e^{-\delta t}V^C(X_t^*, Y_t, \Theta_t^*)) = 0$ as

$$\mathbb{E}(V^C(x,y,\theta)\exp\left(-n(q^*)t\right)H_t) = V^C(x,y,\theta)\exp\left(-n(q^*)t\right) \to 0.$$

3.F The first order differential equation

For convenience, we recall some notations, and introduce some more:

$$\begin{aligned}
m(q) &= \frac{R(1-R)}{b_1}q^2 - \frac{b_3(1-R)}{b_1}q + 1,\\
\ell(q) &= m(q) + \frac{1-R}{b_1}q(1-q) + \frac{(b_2-1)R(1-R)}{b_1}\frac{q}{(1-R)q+R},\\
\varphi(q,n) &= b_1(n-1) + (1-R)(b_3-2R)q + (2-b_2)R(1-R),\\
E(q)^2 &= 4R^2(1-R)^2(b_2-1)(1-q)^2,\\
v(q,n) &= \varphi(q,n) - \operatorname{sgn}(1-R)\sqrt{\varphi(q,n)^2 + E(q)^2},\\
D(q,n) &= 2b_1[(1-R)q+R][n-m(q)] - q\left[v(q,n) - v(q,m(q))\right],\\
A(q,n) &= (\ell(q)-n)\left(2b_1[(1-R)q+R] - b_1 q\left(1 - \operatorname{sgn}(1-R)\frac{\varphi}{\sqrt{\varphi^2+E^2}}\right)\right)\\
&\quad +D(q,n).
\end{aligned} \tag{3.37}$$

We begin with a useful lemma.

Lemma 3.13. *$O(q,n)$ has an alternative expression*

$$O(q,n) = -\frac{(1-R)nD(q,n)}{2R(1-q)[(1-R)q+R]b_1[\ell(q)-n]}. \tag{3.38}$$

Proof. Consider

$$\begin{aligned}
&b_1(\ell(q)-n) + \varphi(q,n)\\
&= R(1-R)q^2 - b_3(1-R)q + b_1 - b_1 n + (1-R)q(1-q) + \frac{(b_2-1)R(1-R)q}{(1-R)q+R}\\
&\qquad + b_1 n - b_1 + b_3(1-R)q + R(1-R)[-2q+2-b_2]\\
&= R(1-R)\left[(1-q)^2 - (b_2-1) + \frac{(b_2-1)q}{(1-R)q+R}\right] + (1-R)q(1-q)\\
&= (1-R)(1-q)[R(1-q)+q] - \frac{(b_2-1)R^2(1-R)}{(1-R)q+R}(1-q).
\end{aligned}$$

Then, noting that $(1-R)q+R = R(1-q)+q$,

$$\begin{aligned}
&b_1[(1-R)q+R](\ell(q)-n)\\
&\quad = (1-R)(1-q)[R(1-q)+q]^2 - R^2(1-R)(b_2-1)(1-q) - \varphi(q,n)[R(1-q)+q],
\end{aligned}$$

and multiplying by $4(1-R)(1-q)$,

$$\begin{aligned}
&4b_1(1-R)(1-q)[(1-R)q+R](\ell(q)-n)\\
&\quad = 4(1-R)^2(1-q)^2[R(1-q)+q]^2 - 4\varphi(q,n)(1-R)(1-q)[R(1-q)+q] + \varphi(q,n)^2\\
&\qquad\quad -\{\mathrm{sgn}(1-R)\}^2\left(\varphi(q,n)^2 + 4R^2(1-R)^2(b_2-1)(1-q)^2\right)\\
&\quad = \{2(1-R)(1-q)[R(1-q)+q]-\varphi(q,n)\}^2 - \{\mathrm{sgn}(1-R)\}^2\left\{\varphi(q,n)^2+E(q)^2\right\}.
\end{aligned}$$

Writing this last expression as the difference of two squares we find

$$\begin{aligned}
&2(1-R)(1-q)[(1-R)q+R] - \varphi(q,n) - \mathrm{sgn}(1-R)\sqrt{\varphi(q,n)^2+E(q)^2}\\
&\quad = \frac{4b_1(1-R)(1-q)[(1-R)q+R](\ell(q)-n)}{2(1-R)(1-q)[R(1-q)+q]-v(q,n)}.
\end{aligned}$$

Then

$$\begin{aligned}
O(q,n) &= \frac{(1-R)n}{R(1-q)} - \frac{2(1-R)^2qn/R}{2(1-R)(1-q)\left[(1-R)q+R\right]-\varphi(q,n)-\mathrm{sgn}(1-R)\sqrt{\varphi(q,n)^2+E(q)^2}}\\
&= \frac{(1-R)n}{R(1-q)}\left\{1-\frac{(1-R)q(1-q)}{b_1(\ell(q)-n)}+\frac{qv(q,n)}{2b_1[(1-R)q+R](\ell(q)-n)}\right\} \qquad (3.39)\\
&= \frac{(1-R)n\left\{2b_1(\ell(q)-n)[(1-R)q+R]-2[(1-R)q+R](1-R)q(1-q)+qv(q,n)\right\}}{2b_1R[(1-R)q+R](1-q)(\ell(q)-n)}\\
&= \frac{(1-R)n\left\{2b_1[(1-R)q+R]\left[(\ell(q)-m(q))-(n-m(q))-\frac{(1-R)q(1-q)}{b_1}\right]+qv(q,n)\right\}}{2b_1R(1-q)[(1-R)q+R](l(q)-n)}.
\end{aligned}$$

The result then follows since

$$2b_1[(1-R)q+R]\left\{\ell(q)-m(q)-\frac{(1-R)q(1-q)}{b_1}\right\} = 2R(1-R)(b_2-1)q = -qv(q,m).$$

$\square$

Proof of Lemma 3.6. (1) Observe that

$$\begin{aligned}
\ell(q)-m(q) &= \frac{1-R}{b_1}q(1-q) + \frac{(b_2-1)R(1-R)}{b_1}\frac{q}{(1-R)q+R}\\
&= \frac{(1-R)q}{b_1[(1-R)q+R]}P(q)
\end{aligned}$$

where $P(q) = Rb_2 + (1-2R)q - (1-R)q^2$. Hence the crossing points of $\ell(q)$ and $m(q)$ away from $q = 0$ are given by the roots of $P(q) = 0$ if such roots exist. Note that $P(-\frac{R}{1-R}) = P(1) = R(b_2 - 1) > 0$, since by assumption, $b_2 > 1$.

If $R < 1$, then since P is inverse-U shaped and $P(1) > 0$ there must be two distinct solutions of the quadratic equation $P(q) = 0$. As $0 < P(1) = P(-R/(1-R))$, we must have $P(q) > 0$ on $q \in [-R/(1-R), 1]$, and the two roots must be found outside this interval. If $R > 1$, the minima of $P(q)$ is given by $q_P := \frac{2R-1}{2(R-1)}$. Note that $1 < q_P < R/(R-1)$, and since $0 < P(1) = P(R/(R-1))$, the root(s) of $P(q) = 0$ must be contained on the interval $(1, R/(R-1))$ if they exist. The desired results can be established easily using these properties of P.

(2) The behaviour at $q = -R/(1-R)$ is only relevant for $R > 1$ so we write this as $q = R/(R-1)$. Note that ℓ explodes at $q = \frac{R}{R-1}$. It is sufficient to check the denominator of $O(q,n)$ is not equal to zero at $q = R/(R-1)$. Direct calculation gives

$$[(1-R)q + R][\ell(q) - n]|_{q=\frac{R}{R-1}} = -\frac{(b_2-1)R^2}{b_1}$$

and hence

$$2R(1-q)[(1-R)q+R]b_1[\ell(q)-n]|_{q=\frac{R}{1-R}} = \frac{2R^3(b_2-1)}{(R-1)} \neq 0. \tag{3.40}$$

(3) The following lemma records some useful identities.

Lemma 3.14.

$$\begin{aligned}
\varphi(q, m(q)) &= R(1-R)\{(1-q)^2 - (b_2-1)\}, \\
\varphi(q, \ell(q)) &= (1-R)(1-q)\left\{(1-R)q + R - \frac{(b_2-1)R^2}{(1-R)q+R}\right\}, \\
\varphi(1, n) &= b_1(n - \ell(1)), \\
v(q, m(q)) &= -2R(1-R)(b_2-1), \\
v(q, \ell(q)) &= \begin{cases} -\frac{2R^2(1-R)(1-q)(b_2-1)}{(1-R)q+R}, & (1-q)[(1-R)q+R] > 0, \\ 2(1-R)(1-q)[(1-R)q+R], & (1-q)[(1-R)q+R] < 0, \end{cases} \\
v(1, n) &= \varphi(1,n) - \mathrm{sgn}(1-R)|\varphi(1,n)|.
\end{aligned}$$

Proof. Most of these identities follow easily on substitution. For $v(q,\ell(q))$ we have

$$\begin{aligned}
v(q,\ell(q)) &= (1-R)(1-q)\left\{(1-R)q+R-\frac{(b_2-1)R^2}{(1-R)q+R}\right\} \\
&\quad - \operatorname{sgn}(1-R)\sqrt{(1-R)^2(1-q)^2\left\{(1-R)q+R+\frac{(b_2-1)R^2}{(1-R)q+R}\right\}^2} \\
&= (1-R)(1-q)\left\{(1-R)q+R-\frac{(b_2-1)R^2}{(1-R)q+R}\right\} \\
&\quad - (1-R)|1-q|\left|(1-R)q+R+\frac{(b_2-1)R^2}{(1-R)q+R}\right|
\end{aligned}$$

which simplifies to give the stated expression. □

Return to the proof of Part (3) of Lemma 3.6. Note that $\operatorname{sgn}(\varphi(1,n)) = \operatorname{sgn}(n-\ell(1))$. Assume we are in the range $(1-R)n < (1-R)\ell(1)$. Then $\operatorname{sgn}(\varphi(1,n)) = -\operatorname{sgn}(1-R)$, $v(1,n) = 2\varphi(1,n)$ and

$$D(1,n) = 2b_1[n-m(1)]-v(1,n)+v(1,m(1)) = 2b_1[n-m(1)]-2b_1[n-\ell(1)]+2b_1[m(1)-\ell(1)] = 0.$$

Further, after some algebra we can show $\frac{\partial}{\partial q}D(q,n)|_{q=1} = -2b_1R(n-m(1))$.

Consider $F(q,n) = \frac{O(q,n)}{n} = -\frac{(1-R)D(q,n)}{2R(1-q)[(1-R)q+R]b_1[\ell(q)-n]}$. Then both the numerator and denominator of F are zero at $q=1$. Nonetheless, we can apply L'Hôpital's rule to calculate $\lim_{q\to 1}\frac{D(q,n)}{1-q}$ to deduce the expression in (3.15).

Now consider $\lim_{n\to\ell(q)} F(q,n)$. Suppose first $0<q<1$. Then

$$D(q,\ell(q)) = 2(1-R)q(1-q)\left\{[(1-R)q+R]+\frac{R^2(b_2-1)}{(1-R)q+R}\right\}$$

which is non-zero and has $\operatorname{sgn}(D(q,\ell(q))) = \operatorname{sgn}(1-R)$. It follows that for $q<1$, and $R<1$, $\lim_{n\uparrow\ell(q)} F(q,n) = -\infty$ and for $q<1$ and $R>1$, $\lim_{n\downarrow\ell(q)} F(q,n) = +\infty$.

Now suppose $q>1$, and if $R>1$ that $(1-R)q+R>0$. Then

$$\begin{aligned}
D(q,\ell(q)) &= 2b_1[(1-R)q+R]\left(\frac{1-R}{b_1}q(1-q)+\frac{(b_2-1)R(1-R)}{b_1}\frac{q}{(1-R)q+R}\right) \\
&\quad - 2(1-R)q(1-q)[(1-R)q+R]-2R(1-R)(b_2-1)q \\
&= 0.
\end{aligned}$$

Then, in order to determine the value of $F(q,\ell(q))$ via L'Hôpital's rule we need

$$\frac{\partial D}{\partial n} = 2b_1[(1-R)q+R]-q\frac{\partial v}{\partial n} = 2b_1[(1-R)q+R]-b_1q\left(1-\frac{\operatorname{sgn}(1-R)\varphi}{\sqrt{\varphi^2+E^2}}\right). \tag{3.41}$$

It follows that

$$\frac{\partial}{\partial n} D(q,n)\bigg|_{n=\ell} = 2b_1[(1-R)q+R]\left[1 - \frac{q[(1-R)q+R]}{[(1-R)q+R]^2 + R^2(b_2-1)}\right]$$

and hence we obtain (3.16).

(4) We prove the results for $R < 1$. The results for $R > 1$ can be obtained similarly, the only issue being that sometimes there is an extra case which arises when $(1-R)q+R$ changes sign.

Note that for fixed q, the ordering of $m(q)$ and $\ell(q)$ is given by Part 1 of Lemma 3.6. The monotonicity of F in n for $q = 1$ can be obtained from (3.15).

If $0 < q < 1$, then since

$$2b_1[(1-R)q+R] - b_1 q\left(1 - \frac{\operatorname{sgn}(1-R)\varphi}{\sqrt{\varphi^2+E^2}}\right) > 2b_1[(1-R)q+R] - 2b_1 q = 2Rb_1(1-q) > 0,$$

we conclude from (3.41) that $D(q,n)$ is increasing in n. Since $D(q, m(q)) = 0$ it follows that $D(q,n) > 0$ for $n > m(q)$ and $D(q,n) < 0$ for $n < m(q)$. Hence, $F(q,n) = 0$ if and only if $n = m(q)$, and we have

$$\begin{aligned}\operatorname{sgn}(F(q,n)) &= -\operatorname{sgn}\left(\frac{D(q,n)}{(1-q)[(1-R)q+R][\ell(q)-n]}\right)\\ &= \operatorname{sgn}\left[(n-m(q))(n-\ell(q))\right].\end{aligned}$$

This gives the desired sign properties of $F(q,n)$ on the range $0 < q < 1$.

Now consider the case $q > 1$. From Part 3 of this proof, we have $D(q, \ell(q)) = 0$. We can compute the second derivative of D with respect to n as

$$\frac{\partial^2 D}{\partial n^2} = \operatorname{sgn}(1-R) b_1^2 q \frac{E^2}{(E^2+\varphi^2)^{3/2}}$$

so that (recall $R < 1$) $D(q,n)$ is convex in n. Since $D(q,m(q)) = D(q,\ell(q)) = 0$, it follows that on the regime of $q > 1$ we must have $D(q,n) < 0$ when n lies between $m(q)$ and $\ell(q)$ and $D(q,n) > 0$ otherwise. Thus $\operatorname{sgn}(D(q,n)) = \operatorname{sgn}[(n-m(q))(n-\ell(q))]$. Then

$$\operatorname{sgn}(F(q,n)) = \operatorname{sgn}\left(\frac{D(q,n)}{\ell(q)-n}\right) = -\operatorname{sgn}(n-m(q)).$$

Finally, note that $F(q,n)$ can be zero only if $n = m(q)$ or $n = \ell(q)$. But for $q > 1$ the limiting expression at $n = \ell(q)$ is given by Part 3 of Lemma 3.6. Hence $F(q,n) = 0$ if and only if $n = m(q)$.

□

The following lemma on further properties of F is key in the proofs of the monotonicity property of Σ and in results on comparative statics:

Lemma 3.15. *For $q \in (0,1]$ and $(1-R)m(q) < (1-R)n < (1-R)\ell(q)$, and for $q > 1$ and $(1-R)m(q) < (1-R)n$, we have $\frac{\partial}{\partial n}F(q,n) \leqslant 0$.*

Proof. Direct computation gives

$$\begin{aligned}
(\ell(q)-n)^2\frac{\partial}{\partial n}\left(\frac{D(q,n)}{\ell(q)-n}\right) &= (\ell(q)-n)\frac{\partial D}{\partial n} + D(q,n) \\
&= (\ell(q)-n)\left(2b_1[(1-R)q+R] - b_1 q\left(1 - \operatorname{sgn}(1-R)\frac{\varphi}{\sqrt{\varphi^2+E^2}}\right)\right) \\
&\quad + D(q,n) \\
&= A(q,n)
\end{aligned}$$

as defined in (3.37). Differentiating A we have

$$\frac{\partial}{\partial n}A(q,n) = \operatorname{sgn}(1-R)\frac{b_1^2 E(q)^2 q(\ell(q)-n)}{(\varphi^2+E(q)^2)^{3/2}}.$$

Hence for $q > 0$ and $R < 1$, $A(q,n)$ is increasing in n for $n < \ell(q)$ and decreasing in n for $n > \ell(q)$. If $R > 1$, then $A(q,n)$ is decreasing in n for $n < \ell(q)$ and increasing in n for $n > \ell(q)$.

Now we calculate the limiting value of $A(q,n)$ as $n \to \pm\infty$. Clearly $\varphi(q,n) \to \pm\infty$ as $n \to \pm\infty$. Then,

$$\lim_{(1-R)n\to+\infty} v(q,n) = \lim_{(1-R)\varphi\to+\infty}\left(\varphi - \operatorname{sgn}(1-R)\sqrt{\varphi^2+E(q)^2}\right) = 0$$

and

$$\lim_{(1-R)n\to+\infty}(\ell(q)-n)\left(1-\operatorname{sgn}(1-R)\frac{\varphi(n,q)}{\sqrt{\varphi(n,q)^2+E(q)^2}}\right) = 0.$$

Observe that

$$\begin{aligned}
A(q,n) = 2b_1[(1-R)q+R](\ell(q)-m(q)) - b_1 q(\ell(q)-n)\left(1-\operatorname{sgn}(1-R)\frac{\varphi}{\sqrt{\varphi^2+E^2}}\right) \\
- qv(q,n) + qv(q,m(q))
\end{aligned}$$

and thus

$$\lim_{(1-R)n\to+\infty} A(q,n) = 2b_1[(1-R)q+R](\ell(q)-m(q)) + qv(q,m(q)) = 2(1-R)[(1-R)q+R]q(1-q).$$

Now we compute the limiting value of $A(q,n)$ as $\operatorname{sgn}(1-R)n \to -\infty$. In this case $v(q,n)$ is no longer converging. But consider

$$\begin{aligned} & b_1 q(\ell(q)-n)\left(1-\operatorname{sgn}(1-R)\frac{\varphi}{\sqrt{\varphi^2+E^2}}\right)+qv(q,n) \\ = \quad & b_1 q\ell(q)+q(\varphi-b_1 n)-\operatorname{sgn}(1-R)\frac{q\varphi}{\sqrt{\varphi^2+E^2}}\left(b_1\ell(q)+(\varphi-b_1 n)+\frac{E^2}{\varphi}\right). \end{aligned}$$

Using the fact that $\varphi - b_1 n = \varphi(q,n) - b_1 n$ is independent of n, we can obtain

$$\begin{aligned} & \lim_{(1-R)n\to-\infty}\left\{b_1 q(\ell(q)-n)\left(1-\operatorname{sgn}(1-R)\frac{\varphi}{\sqrt{\varphi^2+E^2}}\right)+qv(q,n)\right\} \\ = \quad & 2b_1 q\ell(q)-2q\left[b_1-(1-R)(b_3-2R)q-(2-b_2)R(1-R)\right] \end{aligned}$$

and thus

$$\begin{aligned} \lim_{(1-R)n\to-\infty} A(q,n) &= 2b_1[(1-R)q+R](\ell(q)-m(q))+qv(q,m(q))-2b_1 q\ell(q) \\ &\quad +2q\left[b_1-(1-R)(b_3-2R)q-(2-b_2)R(1-R)\right] \\ &= \frac{2R^2(1-R)(b_2-1)q(1-q)}{(1-R)q+R} \end{aligned}$$

after some algebra.

Suppose $R<1$. For $0<q<1$ we have $A(q,n)$ increasing in n for $n<\ell(q)$ and decreasing in n for $n>\ell(q)$. Since on this range of q $\lim_{n\to+\infty} A(q,n) = 2(1-R)[(1-R)q+R]q(1-q) > 0$ and $\lim_{n\to-\infty} A(q,n) = \frac{2R^2(1-R)(b_2-1)q(1-q)}{(1-R)q+R} > 0$, we conclude $A(q,n)>0$ for all n.

If $q>1$ then $A(q,\ell(q)) = D(q,\ell(q)) = 0$. But $A(q,n)$ attains its maximum at $n=\ell(q)$, hence we have $A(q,n)\leqslant 0$ for $q>1$. Putting the cases together, $(1-q)A(q,n)\geq 0$ and $\frac{\partial F}{\partial n}\leq 0$.

Now suppose $R>1$. Suppose $0<q<1$ or $q>R/(R-1)$. Then $A(q,n)$ decreasing in n for $n<\ell(q)$ and increasing in n for $n>\ell(q)$. Since $\lim_{n\to+\infty} A(q,n)<0$ and $\lim_{n\to-\infty} A(q,n)<0$, we conclude $A(q,n)<0$ for all n. If $1<q<\frac{R}{R-1}$ then $A(q,n)$ attains its minimum of zero at $n=\ell(q)$. Hence $A(q,n)\geqslant 0$ for $1<q<\frac{R}{R-1}$. We find $(1-q)[(1-R)q+R]A(q,n)\leq 0$ and again $\frac{\partial F}{\partial n}\leq 0$.

It remains to check the result at $q=1$ and, if $R>1$, $q=R/(R-1)$. At $q=1$ the result follows by considering (3.15). For $R>1$ and $q=R/(R-1)$, we obtain from (3.40) that

$$F\left(\frac{R}{R-1},n\right) = \frac{(R-1)^2 D(\frac{R}{R-1},n)}{2R^3(b_2-1)}.$$

Using (3.41) we have

$$\left.\frac{\partial D}{\partial n}\right|_{q=R/(R-1)} = -\frac{Rb_1}{R-1}\left(1+\frac{\varphi}{\sqrt{\varphi^2+E^2}}\right) < 0$$

and then the monotonicity of $F(\frac{R}{R-1}, n)$ in n follows from the monotonicity of $D(\frac{R}{R-1}, n)$ in n. □

Proof of Lemma 3.7. For any $u \in (0, q_M)$, since $(1-R)n_u(q)$ is decreasing in q and $n'(\zeta(u)) = 0$, $n_u(q)$ can only cross $m(q)$ at some $q \geqslant q_M$. Moreover, for $u \leq q \leq \zeta(u)$, $(1-R)m(u) = (1-R)n_u(u) \geq (1-R)n_u(q) \geq (1-R)n_u(\zeta(u)) = (1-R)m(\zeta(u)) \geq (1-R)m_M$.

Since $n_{q_M}(q_M) = m_M$, we have $\lim_{u\uparrow q_M} m(\zeta(u)) = m_M$ and in turn $\lim_{u\uparrow q_M}\zeta(u) = q_M$. Then $\lim_{u\uparrow q_M}\Sigma(u) = 0$.

Now consider $\Lambda(u) := \ln(1+\Sigma(u)) = \int_u^{\zeta(u)} -\frac{R}{(1-R)q}\frac{O(q,n_u(q))}{n_u(q)}dq$. From the fact that $O(u, n_u(u)) = O(u, m(u)) = 0 = O(\zeta(u), m(\zeta(u))) = O(\zeta(u), n_u(\zeta(u)))$ we have

$$\frac{d\Lambda}{du} = \int_u^{\zeta(u)} -\frac{R}{(1-R)q}\left(\frac{\partial}{\partial n}\frac{O(q,n_u(q))}{n_u(q)}\right)\frac{\partial n_u(q)}{\partial u}dq < 0$$

where we have used Lemma 3.15 and the monotonicity of n to make the conclusion about the sign.

We now show that $\lim_{u\downarrow 0}\Sigma(u) = +\infty$. We assume $R < 1$; the proof for $R > 1$ is similar. Consider a quadratic function $H(x) = (1-R)(m'(0)-x) - R(l'(0)-x)x$. Then trivially $H(m'(0)) > 0$. Choose a constant k such that $m'(0) < k < \alpha < 0$ where α is the negative root of $H(x) = 0$. Then $H(k) > 0$ and equivalently $k < \frac{(1-R)(m'(0)-k)}{R(l'(0)-k)}$. Now let $b(q) = 1 + kq$. It is clear from the definition of D that $D(0,1) = 0$ and then

$$\left.\frac{d}{dq}D(q,1+kq)\right|_{q=0} = \left.\frac{\partial}{\partial q}D(q,n)\right|_{q=0,n=1} + k\left.\frac{\partial}{\partial n}D(q,n)\right|_{q=0,n=1} = -2Rb_1m'(0) + 2Rb_1k$$

and

$$\lim_{q\downarrow 0} O(q, b(q)) = -\frac{(1-R)\frac{d}{dq}D(q,1+kq)|_{q=0}}{2R^2b_1[\ell'(0)-k]} = \frac{(1-R)(m'(0)-k)}{R(l'(0)-k)}.$$

Then for all $\epsilon > 0$, there exists $K_\epsilon \in (0,1)$ such that $O(q,b(q)) > \frac{(1-R)(m'(0)-k)}{R(l'(0)-k)} - \epsilon$ for $q < K_\epsilon$. Choose ϵ such that $0 < \epsilon < \frac{(1-R)(m'(0)-k)}{R(l'(0)-k)} - k$. Then we have $O(q,b(q)) > k$ on $0 < q < K_\epsilon$ and solutions to $n' = O(q,n)$ can only cross $b(q)$ from below. For fixed $u < K_\epsilon$, let $\psi_u = \inf(q \geqslant u : n_u(q) > b(q))$. Then for $u < q < K_\epsilon \wedge \psi_u$, $n_u'(q) = O(q, n_u(q)) >$

$O(q,b(q)) > k$, where we have used the property that $F(q,n) = \frac{O(q,n)}{n}$ is decreasing in n.[5] Moreover, there also exists K_m such that $m'(q) < \frac{1}{2}(m'(0)+k)$ for $q < K_m$. Hence on $u < q < K_\epsilon \wedge \psi_u \wedge K_m$, $n_u'(q) - m'(q) > k - \frac{1}{2}(m'(0)+k) = \frac{1}{2}(k - m'(0)) =: \widehat{k} > 0$ and then $n_u(q) - m(q) > \widehat{k}(q-u)$. On the other hand, for $\psi_u < q < K_\epsilon \wedge K_m$, $m(q) < 1 + \frac{q}{2}(m'(0)+k)$ and hence $n_u(q) - m(q) > (1+kq) - (1 + \frac{q}{2}(m'(0)+k)) = \widehat{k}q > \widehat{k}(q-u)$. We conclude $n_u(q) - m(q) > \widehat{k}(q-u)$ for $u < q < Q := K_\epsilon \wedge K_m$.

Hence, using (3.38)

$$\begin{aligned}
\ln(1+\Sigma(u)) &= \int_u^{\xi(u)} -\frac{R}{(1-R)q}\frac{O(q,n_u(q))}{n_u(q)}dq \\
&> \int_u^Q \frac{2b_1\,[(1-R)q+R]\,(n_u(q)-m(q)) - q\,[v(q,n_u(q)) - v(q,m(q))]}{2q(1-q)[(1-R)q+R]b_1[l(q)-n_u(q)]}dq.
\end{aligned}$$

For the denominator, and for $u < q < Q \leq 1$ we have

$$\begin{aligned}
2q(1-q)[(1-R)q+R]b_1[l(q)-n_u(q)] &< 2q(1-q)[(1-R)q+R]b_1[l(q)-m(q)] \\
&= 2q^2(1-q)\{(1-R)(1-q)[(1-R)q+R] \\
&\qquad + (b_2-1)R(1-R)\} \\
&< 2q^2\{M + (b_2-1)R(1-R)\}
\end{aligned}$$

where $M := \sup_{0<q<1}(1-R)(1-q)[(1-R)q+R]$. For the numerator, note that for $q < \zeta(u)$

$$\begin{aligned}
&v(q,n_u(q)) - v(q,m(q)) \\
&= \varphi(q,n_u(q)) - \varphi(q,m(q)) - \{\sqrt{\varphi(q,n_u(q))^2 + E(q)^2} - \sqrt{\varphi(q,m(q))^2 + E(q)^2}\} \\
&< \varphi(q,n_u(q)) - \varphi(q,m(q)) \\
&= b_1(n_u(q) - m(q)).
\end{aligned}$$

Then,

$$\begin{aligned}
&2b_1[(1-R)q+R](n_u(q)-m(q)) - q\,[v(q,n_u(q)) - v(q,m(q))] \\
&> \{2b_1[(1-R)q+R] - b_1 q\}(n_u(q)-m(q)) \\
&= b_1 L(q)(n_u(q)-m(q))
\end{aligned}$$

[5] In the case of $R > 1$, the fact that $F(q,n)$ is decreasing in n alone is not sufficient to conclude $O(q,n)$ is also decreasing in n since F is positive. But one could directly compute

$$\frac{\partial O}{\partial n} = \frac{R-1}{2R(1-q)[(1-R)q+R]b_1(\ell-n)^2}\left(\ell D(q,n) + n(\ell-n)\frac{\partial D}{\partial n}\right)$$

and check the above expression is negative on $0 < q < 1$ since on this range of q we have $D(q,n) < 0$ for $n < m(q)$ and $\frac{\partial D}{\partial n} > 0$.

where $L(q) := \{2[(1-R)q + R] - q\}$. Since L is linear and $L(0) = 2R > 0$, we can choose to work on a small interval $(0, q_L)$ such that $L(q) > \min(2R, L(q_L)) > 0$. For sufficiently small u such that $u < q_L$, we have $b_1 L(q)(n_u(q) - m(q)) > b_1 \min(2R, L(q_L))\widehat{k}(q-u)$ on $u < q < Q \wedge q_L$.

Putting everything together and setting $\widehat{Q} := Q \wedge q_L \wedge 1$, for $u < \widehat{Q}$ we deduce that

$$\begin{aligned}\ln(1+\Sigma(u)) &> \int_u^{\widehat{Q}} \frac{b_1 \min(2R, L(q_L))\widehat{k}(q-u)}{2q^2[M + (b_2-1)R(1-R)]} dq \\ &= \frac{b_1 \min(2R, L(q_L))\widehat{k}}{2[M + (b_2-1)R(1-R)]}\left(\ln\frac{\widehat{Q}}{u} + \frac{u}{\widehat{Q}} - 1\right).\end{aligned}$$

Letting $u \downarrow 0$ and noting that $\widehat{Q}$ does not depend on u we conclude that $\Sigma(u) \to \infty$.

$\square$

3.G Comparative statics

Proof of Proposition 3.8. (1) Set $\overline{m}(q) = b_1(m(q)-1)$ and similarly $\overline{n}(q) = b_1(n(q)-1)$ and $\overline{\ell}(q) = b_1(\ell(q)-1)$. The idea behind this transformation is that $\overline{m}$ is constructed such that it does not depend on b_1. ℓ has a similar property. The free boundary value problem can be written as to find $(\overline{n}, q_*, q^*)$ such that $\overline{n}' = \overline{O}(q, \overline{n})$ subject to $\overline{n}(q_*) = \overline{m}(q_*)$ and $\overline{n}(q^*) = \overline{m}(q^*)$. Here $\overline{O}(q, \overline{n}) := b_1 O(q, \frac{\overline{n}}{b_1} + 1) = b_1 O(q, n)$.

Note that $\zeta(u) = \inf\{q \geqslant u : (1-R)n_u(q) < (1-R)m(q)\} = \inf\{q \geqslant u : (1-R)\overline{n}_u(q) < (1-R)\overline{m}(q)\}$.

Define $\overline{\varphi}(q, \overline{n}) = \varphi(q, n) = \varphi(q, \frac{\overline{n}}{b_1} + 1)$, $\overline{v}(q, \overline{n}) = v(q, n) = v(q, \frac{\overline{n}}{b_1} + 1)$ and $\overline{D}(q, \overline{n}) = D(q, n) = D(q, \frac{\overline{n}}{b_1} + 1)$. Then, as functions of q and $\overline{n}$, $\overline{\varphi}$, $\overline{v}$ and $\overline{D}$ are all independent of b_1.

We have

$$\overline{O}(q, \overline{n}) = -\frac{(1-R)(\overline{n} + b_1)\overline{D}(q, \overline{n})}{2R(1-q)[(1-R)q + R][\overline{\ell}(q) - \overline{n}]}.$$

By the above remarks the only dependence on b_1 is through the term $(\overline{n} + b_1)$. Further

$$\overline{n}' = (\overline{n} + b_1)\overline{F}(q, \overline{n})$$

where $\overline{F}$ given by

$$\overline{F}(q, \overline{n}) = F(q, n) = -\frac{(1-R)\overline{D}(q, \overline{n})}{2R(1-q)[(1-R)q + R][\overline{\ell}(q) - \overline{n}]}$$

does not depend on b_1. By Lemma 3.15, $\overline{F}$ is decreasing in the second argument.

Let $\widehat{b}_1 > \widetilde{b}_1$ be two positive values of b_1. Define $\widehat{n}_u$ and $\widetilde{n}_u$ the solutions to the initial value problem $\overline{n}'(q) = \overline{O}(q, \overline{n}(q))$ with $\overline{n}(u) = \overline{m}(u)$ under parameters $\widehat{b}_1$ and $\widetilde{b}_1$ respectively. We extend this notation to O, ζ, Σ and (q_*, q^*) in a similar fashion.

If $\overline{n}_u$ is a solution to the initial value problem with $\overline{n}_u(u) = \overline{m}(u)$ we must have $(1-R)\overline{O}(q, \overline{n}_u(q)) < 0$ and hence $(1-R)\overline{O}$ is decreasing in b_1. Then $(1-R)\widehat{n}_u$ cannot upcross $(1-R)\widetilde{n}_u$ and since $(1-R)\widehat{n}_u'(u) = (1-R)\widehat{O}(u, \widehat{n}_u(u)) < (1-R)\widetilde{O}(u, \widetilde{n}_u(u)) = (1-R)\widetilde{n}_u'(u)$, we must have $(1-R)\widehat{n}_u(q) < (1-R)\widetilde{n}_u(q)$ at least up to $q = \widehat{\zeta}(u) \wedge \widetilde{\zeta}(u)$. From this we conclude $\widehat{\zeta}(u) < \widetilde{\zeta}(u)$. On the other hand, $\overline{F}(q, \overline{n})$ depends on b_1 only through $\overline{n}$. It follows that

$$
\begin{aligned}
-\ln(1+\overline{\Sigma}(u)) &= \int_u^{\overline{\zeta}(u)} \frac{R}{q(1-R)} \frac{O(q, \frac{\overline{n}_u(q)}{b_1} + 1)}{\frac{\overline{n}_u(q)}{b_1} + 1} dq \\
&= \int_u^{\overline{\zeta}(u)} \frac{R}{q(1-R)} \frac{\overline{O}(q, \overline{n}_u(q))}{\overline{n}_u(q) + b_1} dq = \int_u^{\overline{\zeta}(u)} \frac{R}{q(1-R)} \overline{F}(q, \overline{n}_u(q)) dq.
\end{aligned}
$$

But, by the monotonity of $\overline{n}_u$ and $\overline{\zeta}$ in b_1

$$
\int_u^{\widehat{\zeta}(u)} \frac{R}{q(1-R)} \overline{F}(q, \widehat{n}_u(q)) dq > \int_u^{\widehat{\zeta}(u)} \frac{R}{q(1-R)} \overline{F}(q, \widetilde{n}_u(q)) dq > \int_u^{\widetilde{\zeta}(u)} \frac{R}{q(1-R)} \overline{F}(q, \widetilde{n}_u(q)) dq
$$

where we use $(1-R)\overline{F}(q, n) < 0$ and and the fact that $\overline{F}$ is decreasing in $\overline{n}$ over the relevant range. We conclude that $\ln(1+\widehat{\Sigma}(u)) < \ln(1+\widetilde{\Sigma}(u))$ and hence $\widehat{q}_* = \widehat{\Sigma}^{-1}(\xi) < \widetilde{\Sigma}^{-1}(\xi) = \widetilde{q}_*$.

To prove the monotonicity of the sale boundary q^*, one can parameterise the family of solutions via its right boundary point $(n_v(\cdot), \varsigma(v), v)$. See Hobson et al. (2016) for the use of a similar idea.

(2) Now we consider the monotonicity of the limits of the no-transaction wedge in b_3. We use a different transformation and comparison result. Set $a(q) = n(q) - m(q)$. Then the original free boundary value problem becomes to solve $a'(q) = \underline{O}(q, a(q))$ subject to boundary conditions $a(q_*) = a(q^*) = 0$ where

$$
\underline{O}(q, a) = -\frac{(1-R)(a + m(q))D(q, a + m(q))}{2R(1-q)[(1-R)q + R]b_1[\ell(q) - m(q) - a]} - \frac{2R(1-R)}{b_1} q + \frac{b_3(1-R)}{b_1}.
$$

Observe that $b_1[\ell(q) - m(q)] = (1-R)q(1-q) + (b_2 - 1)R(1-R)\frac{q}{(1-R)q+R}$ does not depend on b_3. Further,

$$
\varphi(q, a + m(q)) = b_1 a + \varphi(q, m(q)) = b_1 a + R(1-R)\{(1-q)^2 - (b_2 - 1)\}
$$

and $v(q, m(q)) = -2R(1-R)(b_2 - 1)$ are both independent of b_3. Hence $D(q, a + m(q)) = 2b_1[(1-R)q + R]a - q[v(q, a + m(q)) - v(q, m(q))]$ and

$$
\frac{O(q, a + m(q))}{a + m(q)} = \frac{(1-R)D(q, a + m(q))}{2R(1-q)[(1-R)q + R]b_1[\ell(q) - m(q) - a]}
$$

are independent of b_3. Recall we are assuming $R<1$. Then $O(q,n)\le 0$ over the relevant range and

$$\begin{aligned}\frac{\partial \underline{O}}{\partial b_3}(q,a) &= -\frac{(1-R)D(q,a+m(q))}{2R(1-q)[(1-R)q+R]b_1[\ell(q)-m(q)-a]}\frac{\partial m}{\partial b_3}+\frac{1-R}{b_1}\\ &= -\frac{O(q,a+m(q))}{a+m(q)}\times\frac{1-R}{b_1}q+\frac{1-R}{b_1}>0.\end{aligned}$$

Suppose $\widehat{b}_3>\widetilde{b}_3$. Using similar ideas in Part 1 of the proof we can deduce $\widehat{\zeta}(u)>\widetilde{\zeta}(u)$ and $\widehat{a}_u(q)>\widetilde{a}_u(q)$ for $q<\widetilde{\zeta}(u)$. Hence, using the fact that $\frac{O(q,a+m(q))}{a+m(q)}$ does not depend on b_3

$$\begin{aligned}\ln(1+\widehat{\Sigma}(u)) &= \int_u^{\widehat{\zeta}(u)}\left(-\frac{R}{q(1-R)}\frac{O(q,\widehat{n}_u(q))}{\widehat{n}_u(q)}\right)dq\\ &= \int_u^{\widehat{\zeta}(u)}\left(-\frac{R}{q(1-R)}\frac{O(q,\widehat{a}_u(q)+m(q))}{\widehat{a}_u(q)+m(q)}\right)dq\\ &> \int_u^{\widetilde{\zeta}(u)}\left(-\frac{R}{q(1-R)}\frac{O(q,\widehat{a}_u(q)+m(q))}{\widehat{a}_u(q)+m(q)}\right)dq\\ &> \int_u^{\widetilde{\zeta}(u)}\left(-\frac{R}{q(1-R)}\frac{O(q,\widetilde{a}_u(q)+m(q))}{\widetilde{a}_u(q)+m(q)}\right)dq\\ &= \ln(1+\widetilde{\Sigma}(u))\end{aligned}$$

where we use the monotonicity of $\zeta(u)$ and the property that $\frac{O(q,n)}{n}$ is decreasing in n and hence $\frac{O(q,a+m(q))}{a+m(q)}$ is decreasing in a. Thus $\widehat{q}_*=\widehat{\Sigma}^{-1}(\xi)>\widetilde{\Sigma}^{-1}(\xi)=\widetilde{q}_*$. The monotonicity property of the sale boundary can be proved in a similar fashion by parameterising the family of solutions with their right boundary points.

□

Proof of Theorem 3.11. (1) We write out the proof assuming $R<1$. The case $R>1$ follows similarly.

We use (3.7) to compute

$$\begin{aligned}\frac{\partial}{\partial b_1}O(q,n;b_1) &= -\frac{2(1-R)^2qn/R}{\{2(1-R)(1-q)[(1-R)q+R]-\varphi(q,n)-\sqrt{\varphi(q,n)^2+E(q)^2}\}^2}\\ &\quad\times\left(1+\frac{\varphi(q,n)}{\sqrt{\varphi(q,n)^2+E(q)^2}}\right)\frac{\partial\varphi}{\partial b_1}\end{aligned}$$

and hence, for $q>0$, $\operatorname{sgn}\left(\frac{\partial}{\partial b_1}O(q,n;b_1)\right)=-\operatorname{sgn}\left(\frac{\partial\varphi}{\partial b_1}\right)=-\operatorname{sgn}(n-1)=+1$, since $n(\cdot)$ is bounded above by 1.

Further, $\overline{m}(q):=b_1(m(q;b_1)-1)$ is independent of b_1 and from this we deduce $\frac{\partial}{\partial b_1}m(q;b_1)=-\frac{m(q;b_1)-1}{b_1}$ and hence over the continuation region $q\in[q_*,q^*]$ we have

$\operatorname{sgn}\left(\frac{\partial}{\partial b_1}m(q;b_1)\right) = -\operatorname{sgn}(m(q;b_1)-1) = +1$. Using the signs of $\frac{\partial}{\partial b_1}O(q,n;b_1)|_{n=n(q)}$ and $\frac{\partial}{\partial b_1}m(q;b_1)$ together with the fact that q_* is decreasing in b_1, we conclude $n(\,\cdot\,;b_1)$ is increasing in b_1. If we extend the domain of definition of n to $(0,\infty)$ by setting $n(q) = n(q_*)$ for $q < q_*$ and $n(q) = n(q^*)$ for $q > q^*$ then we have $n(\,\cdot\,;b_1)$ being increasing in b_1 on $(0,\infty)$.

Starting from the fact that $n(q;b_1)$ is increasing in b_1, we can deduce that each of $-(1-q)N(q;b_1)$, $hW(h;b_1)$, $w(h;b_1)$ and $(1-p)h'(p;b_1)$ is increasing in b_1. Then for $\widehat{b}_1 > \widetilde{b}_1$ (and using the overscripts to label the functions and parameters under the corresponding choice of b_1), we have

$$\operatorname{sgn}(1-p)\widehat{h}'(p) > \operatorname{sgn}(1-p)\widetilde{h}'(p). \tag{3.42}$$

Recall that $G(p) = n(q_*)^{-R}(1+\lambda p)^{1-R}$ and $G(p) = n(q^*)^{-R}(1-\gamma p)^{1-R}$ on the purchase and sale region respectively. Using the monotonicity of n in b_1 we conclude $\widehat{G}(p) < \widetilde{G}(p)$ over $p \in (0,\widehat{p}_*) \cup (\widetilde{p}^*, 1/\gamma)$.

Suppose $G(p;b_1)$ is not decreasing in b_1. Then since G is continuous, $\widehat{G}(p)$ must cross $\widetilde{G}(p)$ at least twice, with the first cross being an upcross and the last cross being a downcross. Denote the p-coordinate of the first upcross and last downcross by k_u and k_d respectively.

Away from $p = 1$, (3.42) implies that $\widehat{G}(p)$ cannot downcross $\widetilde{G}(p)$. Then the only possibility is that there are precisely two crossings with $0 < k_u < k_d = 1$. But if $k_d = 1$ such that $K := \widehat{G}(1) = \widetilde{G}(1)$, the relationship $\frac{1}{G(1)}\left(G(1) - \frac{G'(1)}{1-R}\right)^{1-1/R} = n(1)$ gives

$$\widehat{G}'(1) = (1-R)\left(K - (K\widehat{n}(1))^{-R/(1-R)}\right) > (1-R)\left(K - (K\widetilde{n}(1))^{-R/(1-R)}\right) = \widetilde{G}'(1)$$

contradicting the hypothesis that $k_d = 1$ is a downcross.

(2) For $R < 1$ a similar argument to the above can be applied if we can show that $n(\,\cdot\,;b_3)$ is decreasing in b_3. But this follows immediately as $\operatorname{sgn}\left(\frac{\partial}{\partial b_3}O(q,n;b_3)\right) = -\operatorname{sgn}\left(\frac{\partial \varphi}{\partial b_3}\right) = -1 = \operatorname{sgn}\left(\frac{\partial}{\partial b_3}m(q;b_3)\right)$ and q_* is increasing in b_3.

If $R > 1$ we cannot use this argument. However, the monotonicity of the value function in b_3, and hence the monotonicity of $\mathcal{C}$ can be proved by a comparison argument.[6] Consider a pair of models, the only difference being that in the first model Y has drift $\tilde{\alpha}$, whereas in the second model Y has drift $\hat{\alpha}$ where $\hat{\alpha} > \tilde{\alpha}$. Write $\epsilon = \hat{\alpha} - \tilde{\alpha} > 0$. Suppose that parameters are such that Standing Assumption 3.1 holds in the first model; then necessarily

[6] The value function only depends on R and the auxiliary parameters, so when comparing two models which differ only through b_3 we may equivalently compare models which differ in α.

Standing Assumption 3.1 holds in the second model. Let $(\tilde{Y}, \hat{Y}) = (\tilde{Y}_t, \hat{Y}_t)_{t\geq 0}$ be given by

$$(\tilde{Y}_t, \hat{Y}_t) = (ye^{\eta W_t + (\tilde{\alpha} - \frac{\eta^2}{2})t}, ye^{\eta W_t + (\hat{\alpha} - \frac{\eta^2}{2})t})$$

so that $\hat{Y}_t = e^{\epsilon t}\tilde{Y}_t$. Let $(\tilde{C}, \tilde{\Pi}, \tilde{\Theta} = \theta + \tilde{\Phi} - \tilde{\Psi})$ be an admissible strategy for an agent in the first model. Suppose $\tilde{\Theta}$ is non-negative, and note that the optimal strategy has this property, even if the initial endowment in the illiquid asset is negative, since in that case there is an initial transaction into the no-transaction wedge which is contained in the half-plane $\theta \geq 0$. We may assume we start in the no-transaction region. Then $\tilde{X}_0 = x$ and $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ solves

$$d\tilde{X}_t = r(\tilde{X}_t - \tilde{\Pi}_t)dt + \frac{\tilde{\Pi}_t}{S_t}dS_t - \tilde{C}_t dt - \tilde{Y}_t(1+\lambda)d\tilde{\Phi}_t + \tilde{Y}_t(1-\gamma)d\tilde{\Psi}_t.$$

Define the absolutely continuous, increasing process κ by $\kappa_t = \int_0^t \left\{ d\tilde{\Phi}_s \wedge (d\tilde{\Psi}_s + \epsilon\tilde{\Theta}_s ds) \right\}$ and set

$$\begin{aligned}
\hat{\Pi}_t &= \tilde{\Pi}_t, \\
\hat{\Theta}_t &= \theta + \hat{\Phi}_t - \hat{\Psi}_t, \\
\hat{C}_t &= \tilde{C}_t + (\lambda+\gamma)\tilde{Y}_t d\kappa_t + (1-\gamma)\epsilon\tilde{\Theta}_t\tilde{Y}_t dt, \\
\hat{\Phi}_t &= \int_0^t e^{-\epsilon s}\left(d\tilde{\Phi}_s - d\kappa_s\right), \\
\hat{\Psi}_t &= \int_0^t e^{-\epsilon s}\left(d\tilde{\Psi}_s + \epsilon\tilde{\Theta}_s ds - d\kappa_s\right).
\end{aligned}$$

Then

$$d\hat{\Theta}_t = d\hat{\Phi}_t - d\hat{\Psi}_t = e^{-\epsilon t}d\tilde{\Theta}_t - \epsilon e^{-\epsilon t}\tilde{\Theta}_t dt = d(e^{-\epsilon t}\tilde{\Theta}_t)$$

which gives $\hat{\Theta}_t = \tilde{\Theta}_t e^{-\epsilon t}$ and in turn $\hat{\Theta}_t\hat{Y}_t = \tilde{\Theta}_t\tilde{Y}_t$. Then the corresponding wealth process solves

$$\begin{aligned}
d\hat{X}_t &= r(\hat{X}_t - \hat{\Pi}_t)dt + \frac{\hat{\Pi}_t}{S_t}dS_t - \hat{Y}_t(1+\lambda)d\hat{\Phi}_t + \hat{Y}_t(1-\gamma)d\hat{\Psi}_t - \hat{C}_t dt \\
&= r(\hat{X}_t - \tilde{\Pi}_t)dt + \frac{\tilde{\Pi}_t}{S_t}dS_t - \hat{Y}_t e^{-\epsilon t}(1+\lambda)[d\tilde{\Phi}_t - d\kappa_t] + \hat{Y}_t e^{-\epsilon t}(1-\gamma)[d\tilde{\Psi}_t + \epsilon\tilde{\Theta}_t dt - d\kappa_t] \\
&\quad - \tilde{C}_t dt - (1-\gamma)\epsilon\tilde{\Theta}_t\tilde{Y}_t dt - (\lambda+\gamma)\tilde{Y}_t d\kappa_t \\
&= r(\hat{X}_t - \tilde{\Pi}_t)dt + \frac{\tilde{\Pi}_t}{S_t}dS_t - \tilde{C}_t dt - \tilde{Y}_t(1+\lambda)d\tilde{\Phi}_t + \tilde{Y}_t(1-\gamma)d\tilde{\Psi}_t.
\end{aligned}$$

If $\hat{X}_0 = x = \tilde{X}_0$ then $\hat{X}$ solves the same equation as $\tilde{X}$ and $\hat{X}_t = \tilde{X}_t \geq 0$. Then, for any admissible strategy in the first model for which $(\Theta_t)_{t\geq 0}$ is positive, including the optimal strategy in this model, there is a corresponding admissible strategy in the second model with strictly larger consumption at all future times. Hence the value function is strictly greater in the second model. $\square$

3.H The consistency condition on transaction costs

Fix positive constant $\epsilon > 0$ and define $\delta_1(\epsilon) := 1 - W(h(1-\epsilon)) > 0$ and $\delta_2(\epsilon) := W(h(1+\epsilon)) - 1 > 0$. Then for

$$o(\epsilon, \delta_1, \delta_2) := \ln\left(\frac{1-\epsilon}{1+\epsilon}\right) - \ln\left(\frac{\delta_2(1-\delta_1)}{\delta_1(1+\delta_2)}\right),$$

we have for $p_* < 1 < p^*$

$$\begin{aligned}
&\ln(1+\xi) + o(\epsilon, \delta_1(\epsilon), \delta_2(\epsilon)) \\
&= \left[\int_{p_*}^{1-\epsilon} \frac{dp}{p(1-p)} + \int_{1+\epsilon}^{p^*} \frac{dp}{p(1-p)}\right] - \left[\int_{q_*}^{1-\delta_1(\epsilon)} \frac{dq}{q(1-q)} + \int_{1+\delta_2(\epsilon)}^{q^*} \frac{dq}{q(1-q)}\right] \\
&= \left[\int_{h_*}^{h(1-\epsilon)} \frac{dh}{w(h)} + \int_{h(1+\epsilon)}^{h^*} \frac{dh}{w(h)}\right] - \left[\int_{q_*}^{1-\delta_1(\epsilon)} \frac{dq}{q(1-q)} - \int_{1+\delta_2(\epsilon)}^{q^*} \frac{dq}{q(1-q)}\right] \\
&= \left[\int_{q_*}^{W(h(1-\epsilon))} \frac{N'(q)dq}{(1-R)qN(q)} + \int_{W(h(1+\epsilon))}^{q^*} \frac{N'(q)dq}{(1-R)qN(q)}\right] \\
&\quad - \left[\int_{q_*}^{1-\delta_1(\epsilon)} \frac{dq}{q(1-q)} - \int_{1+\delta_2(\epsilon)}^{q^*} \frac{dq}{q(1-q)}\right] \\
&= \int_{q_*}^{1-\delta_1(\epsilon)} \left(-\frac{R}{q(1-R)}\frac{O(q,n(q))}{n(q)}\right) dq + \int_{1+\delta_2(\epsilon)}^{q^*} \left(-\frac{R}{q(1-R)}\frac{O(q,n(q))}{n(q)}\right) dq.
\end{aligned}$$

On sending $\epsilon \downarrow 0$, we have $\delta_1(\epsilon) \downarrow 0$ and $\delta_2(\epsilon) \downarrow 0$ and thus

$$\int_{q_*}^{q^*} \left(-\frac{R}{q(1-R)}\frac{O(q,n(q))}{n(q)}\right) dq = \ln(1+\xi) + \lim_{\epsilon\downarrow 0} o(\epsilon, \delta_1(\epsilon), \delta_2(\epsilon)).$$

Now,

$$\frac{\delta_2(\epsilon)}{\delta_1(\epsilon)} = \frac{W(h(1+\epsilon)) - 1}{1 - W(h(1-\epsilon))} = \frac{W(h(1+\epsilon)) - 1}{\epsilon} \frac{\epsilon}{1 - W(h(1-\epsilon))}.$$

But

$$\frac{1 - W(h(1-\epsilon))}{\epsilon} = \frac{(1-R)h(1-\epsilon) - \epsilon(1-\epsilon)h'(1-\epsilon)}{(1-R)\epsilon h(1-\epsilon)} = 1 - \frac{(1-\epsilon)G'(1-\epsilon)}{(1-R)G(1-\epsilon)}$$

and thus $\lim_{\epsilon\downarrow 0} \frac{1-W(h(1-\epsilon))}{\epsilon} = 1 - \frac{G'(1)}{(1-R)G(1)}$. Similarly, we have $\lim_{\epsilon\downarrow 0} \frac{W(h(1+\epsilon))-1}{\epsilon} = 1 - \frac{G'(1)}{(1-R)G(1)}$. Hence $\lim_{\epsilon\downarrow 0} o(\epsilon, \delta_1(\epsilon), \delta_2(\epsilon)) = 0$ and (3.11) holds. In case either $p_* = 1$ or $p^* = 1$, a similar argument can be used to show that (3.11) is still valid.

Bibliography

Agranov, M. and Ortoleva, P. (2015). Stochastic choice and preferences for randomization. *Available at SSRN 2644784*.

Akian, M., Menaldi, J., and Sulem, A. (1995). Multi-asset portfolio selection problem with transaction costs. *Mathematics and Computers in Simulation*, 38(1-3):163–172.

An, L. (2016). Asset pricing when traders sell extreme winners and losers. *Review of Financial Studies*, 29(3):823–861.

Azéma, J. and Yor, M. (1979). Une solution simple au probleme de skorokhod. In *Séminaire de Probabilités XIII*, pages 90–115. Springer.

Baele, L., Driessen, J., Londono, J. M., and Spalt, O. G. (2014). Cumulative prospect theory and the variance premium. *Available at SSRN 2517657*.

Barber, B. M. and Odean, T. (2013). The behavior of individual investors. In *Handbook of the Economics of Finance*, pages 1533–1570. North Holland.

Barberis, N. (2012). A model of casino gambling. *Management Science*, 58(1):35–51.

Barberis, N. (2013). The psychology of tail events: Progress and challenges. *The American Economic Review*, 103(3):611–616.

Barberis, N. and Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *The American Economic Review*, 98(5):2066–2100.

Barberis, N. and Xiong, W. (2009). What drives the disposition effect? an analysis of a long-standing preference-based explanation. *The Journal of Finance*, 64(2):751–784.

Barberis, N. and Xiong, W. (2012). Realization utility. *Journal of Financial Economics*, 104(2):251–271.

Ben-David, I. and Hirshleifer, D. (2012). Are investors really reluctant to realize their losses? trading responses to past returns and the disposition effect. *Review of Financial Studies*, 25(8):2485–2532.

Benartzi, S. and Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, 110(1):73–92.

Bertoin, J., Le Jan, Y., et al. (1992). Representation of measures by balayage from a regular recurrent point. *The Annals of Probability*, 20(1):538–548.

Birru, J. (2015). Confusion of confusions: a test of the disposition effect and momentum. *Review of Financial Studies*, 28(7):1849–1873.

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2009). Fight or flight? portfolio rebalancing by individual investors. *The Quarterly Journal of Economics*, 124(1):301–348.

Camerer, C. F. and Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2):167–196.

Carassus, L. and Rasonyi, M. (2015). On optimal investment for a behavioral investor in multiperiod incomplete market models. *Mathematical Finance*, 25(1):115–153.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728.

Cerreia-Vioglio, S., Dillenberger, D., Ortoleva, P., and Riella, G. (2013). Deliberately stochastic: Random choice and preferences for hedging. Working paper.

Chateauneuf, A., Eichberger, J., and Grant, S. (2007). Choice under uncertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory*, 137(1):538–567.

Chen, X. and Dai, M. (2013). Characterisation of optimal strategy for multi-asset investment and consumption with transaction costs. *SIAM Journal on Financial Mathematics*, 4(1):857–883.

Choi, J. H., Sirbu, M., and Zitkovic, G. (2013). Shadow prices and well-posedness in the problem of optimal investment and consumption with transaction costs. *SIAM Journal on Control and Optimization*, 51(6):4414–4449.

Collings, P. and Haussmann, U. G. (1999). Optimal portfolio selection with transaction costs. In *Control of Distributed Parameter and Stochastic Systems*, pages 189–197. Springer.

Conrad, J., Dittmar, R. F., and Ghysels, E. (2013). Ex ante skewness and expected stock returns. *The Journal of Finance*, 68(1):85–124.

Davis, M. H. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):676–713.

De Giorgi, E. G. and Legg, S. (2012). Dynamic portfolio choice and asset pricing with narrow framing and probability weighting. *Journal of Economic Dynamics and Control*, 36(7):951–972.

Dhar, R. and Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 52(5):726–740.

Dwenger, N., Kübler, D., and Weizsäcker, G. (2013). Flipping a coin: Theory and evidence. *Available at SSRN 2353282*.

Ebert, S. and Hilpert, C. (2015). Skewness preference and the popularity of technical analysis. *Available at SSRN 2354962*.

Ebert, S. and Strack, P. (2015). Until the bitter end: on prospect theory in a dynamic context. *The American Economic Review*, 105(4):1618–1633.

Evans, J., Henderson, V., and Hobson, D. (2008). Optimal timing for an indivisible asset sale. *Mathematical Finance*, 18(4):545–567.

Feng, L. and Seasholes, M. S. (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance*, 9(3):305–351.

Fischbacher, U., Hoffmann, G., and Schudy, S. (2015). The causal effect of stop-loss and take-gain orders on the disposition effect. Working paper.

Genesove, D. and Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, 116(4):1233–1260.

Goldstein, W. M. and Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2):236–254.

Green, T. C. and Hwang, B. H. (2012). Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science*, 58(2):432–444.

Grinblatt, M. and Keloharju, M. (2001). What makes investors trade? *The Journal of Finance*, 56(2):589–616.

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

He, X. D., Hu, S., Obłój, J., and Zhou, X. Y. (2016). Path-dependent and randomized strategies in barberis' casino gambling model. *Available at SSRN 2534936*.

Henderson, V. (2012). Prospect theory, liquidation, and the disposition effect. *Management Science*, 58(2):445–460.

Herczegh, A. and Prokaj, V. (2015). Shadow price in the power utility case. *The Annals of Applied Probability*, 25(5):2671–2707.

Hinkley, D. V. (1975). On power transformations to symmetry. *Biometrika*, 62(1):101–111.

Hobson, D., Tse, A. S. L., and Zhu, Y. (2016). Optimal consumption and investment under transaction costs. Working paper.

Ingersoll, J. E. and Jin, L. J. (2013). Realization utility with reference-dependent preferences. *Review of Financial Studies*, 26(3):723–767.

Jin, L. and Scherbina, A. (2011). Inheriting losers. *Review of Financial Studies*, 24(3):786–820.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kallsen, J. and Muhle-Karbe, J. (2010). On using shadow prices in portfolio optimization with transaction costs. *The Annals of Applied Probability*, 20(4):1341–1358.

Kaminski, K. M. and Lo, A. W. (2014). When do stop-loss rules stop losses? *Journal of Financial Markets*, 18:234–254.

Kaustia, M. (2010). Prospect theory and the disposition effect. *Journal of Financial and Quantitative Analysis*, 45(3):791–812.

Köbberling, V. and Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122(1):119–131.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Kothiyal, A., Spinu, V., and Wakker, P. (2011). Prospect theory for continuous distributions: A preference foundation. *Journal of Risk and Uncertainty*, 42(3):195–210.

Kyle, A. S., Ou-Yang, H., and Xiong, W. (2006). Prospect theory and liquidation decisions. *Journal of Economic Theory*, 129(1):273–288.

Li, Y. and Yang, L. (2013). Prospect theory, the disposition effect, and asset prices. *Journal of Financial Economics*, 107(3):715–739.

Liu, H. (2004). Optimal consumption and investment with transaction costs and multiple risky assets. *The Journal of Finance*, 59(1):289–338.

Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *The Economic Journal*, 95(379):575–594.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, 27(4):1622–1668.

Magill, M. and Constantinides, G. (1976). Portfolio selection with transaction costs. *Journal of Economic Theory*, 13(2):245–263.

Magnani, J. (2015). Testing for the disposition effect on optimal stopping decisions. *The American Economic Review*, 105(5):371–375.

Magnani, J. (2016). Reference-dependent preferences and optimal stopping decisions: a laboratory investigation. Working paper.

Meng, J. and Weng, X. (2016). Can prospect theory explain the disposition effect? a new perspective on reference points. *Available at SSRN 1851883.*

Merton, R. (1969). Life portfolio selection under uncertainty: the continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257.

Muthuraman, K. and Kumar, S. (2006). Multi-dimensional portfolio optimisation with proportional transaction costs. *Mathematical Finance*, 16(2):301–335.

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798.

Polkovnichenko, V. (2005). Household portfolio diversification: A case for rank-dependent preferences. *Review of Financial studies*, 18(4):1467–1502.

Polkovnichenko, V. and Zhao, F. (2013). Probability weighting functions implied in options prices. *Journal of Financial Economics*, 107(3):580–609.

Possamaï, D., Soner, H. M., and Touzi, N. (2015). Homogenization and asymptotics for small transaction costs: the multidimensional case. *Communications in Partial Differential Equations*, 40(11):2005–2046.

Poteshman, A. M. and Serbin, V. (2003). Clearly irrational financial market behavior: Evidence from the early exercise of exchange traded stock options. *The Journal of Finance*, 58(1):37–70.

Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3):497–527.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Raiffa, H. (1961). Risk, ambiguity, and the savage axioms: comment. *The Quarterly Journal of Economics*, 75(4):690–694.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*. Springer-Verlag, Berlin.

Saito, K. (2015). Preferences for flexibility and randomization under uncertainty. *The American Economic Review*, 105(3):1246–1271.

Schneider, C. and Spalt, O. (2016). Conglomerate investment, skewness, and the ceo long-shot bias. *The Journal of Finance*, 71(2):635–672.

Seru, A., Shumway, T., and Stoffman, N. (2010). Learning by trading. *Review of Financial Studies*, 23(2):705–739.

Shefrin, H. and Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3):777–790.

Shreve, S. and Soner, H. (1994). Optimal investment and consumption with transaction costs. *The Annals of Applied Probability*, 4(3):609–692.

Spalt, O. G. (2013). Probability weighting and employee stock options. *Journal of Financial and Quantitative analysis*, 48(4):1085–1118.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31–48.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Viefers, P. and Strack, P. (2014). Too proud to stop: Regret in dynamic decisions. *Available at SSRN 2473538.*

Wakker, P. (1994). Separating marginal utility and probabilistic risk aversion. *Theory and Decision*, 36(1):1–44.

Wakker, P. (2010). *Prospect theory: For risk and ambiguity.* Cambridge university press.

Webb, C. S. (2015). Piecewise additivity for non-expected utility. *Economic Theory*, 60(2):371–392.

Weber, M. and Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33(2):167–184.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42(12):1676–1690.

Xu, Z. Q. and Zhou, X. Y. (2013). Optimal stopping under probability distortion. *The Annals of Applied Probability*, 23(1):251–282.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1):95–115.