MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 411–441

DYNAMIC COHERENT ACCEPTABILITY INDICES AND THEIR APPLICATIONS TO FINANCE

TOMASZ R. BIELECKI, IGOR CIALENCO, AND ZHAO ZHANG

Illinois Institute of Technology

In this paper, we present a theoretical framework for studying coherent acceptability indices (CAIs) in a dynamic setup. We study dynamic CAIs (DCAIs) and dynamic coherent risk measures (DCRMs), and we establish a duality between them. We derive a representation theorem for DCRMs in terms of a so-called dynamically consistent sequence of sets of probability measures. Based on these results, we give a specific construction of DCAIs. We also provide examples of DCAIs, both abstract and also some that generalize selected classical financial measures of portfolio performance.

KEY WORDS: dynamic coherent acceptability index, dynamic measures of performance, dynamic coherent risk measures, dynamically consistent sequence of sets of probability measures, time consistency, dynamic GLR, dynamic RAROC.

1. INTRODUCTION

Individual and institutional investors are typically concerned with finding satisfactory balance between *reward and risk* associated with an investment process. Various measures have been developed to quantify this balance. Such measures are typically referred to as *performance measures* or *measures of performance* (MOP). Recently, Cherny and Madan (2009) originated an effort to provide a mathematical framework to study these measures in a unified way. This paper contributes to this effort.

One of the most popular MOPs is the Sharpe ratio (SR) introduced by Sharpe (1966). SR is expressed as a *ratio* of expected excess return to standard deviation, and thus in financial applications it measures expected excess return of a portfolio in units of portfolio's standard deviation. SR has been used as a classical tool to rank portfolios according to their "reward-to-risk" characteristics.

Using standard deviation to quantify risk is considered to be the major drawback of SR. The reason of course is that positive returns also contribute to this measure of risk. To eliminate this unwanted feature other ratio-types MOPs were proposed, such as Sortino ratio (SOR) (Sortino and Price 1994) and gain loss ratio (GLR) (Bernardo and Ledoit 2000). These MOPs focus on downside risk only. A popular generalization of SR is provided by the risk adjusted return on capital (RAROC), which is constructed as a ratio of mean excess return to some selected measure of risk.

All the MOPs mentioned above share some common desirable features: they are unitless, they are increasing functions of reward and decreasing functions of risk; moreover,

TRB and IC acknowledge support from the NSF grant DMS-0908099. The authors would like to thank the anonymous referees and the editors for their helpful comments and suggestions which improved greatly the final manuscript.

Manuscript received October 2010; final revision received October 2011.

Address correspondence to Igor Cialenco, Department of Applied Mathematics, Illinois Institute of Technology, 10 West 32nd Str, Bld E1, Room 208, Chicago, IL 60616-3793; e-mail: igor@math.iit.edu.

DOI: 10.1111/j.1467-9965.2012.00524.x

according to these MOPs diversification of a portfolio improves its performance. This observation prompts a natural desire to study MOPs in a unified mathematical framework.[1] As already mentioned, such a study was recently originated by Cherny and Madan (2009). We recall the main results of that paper in Section 2. The study of Cherny and Madan (2009) was done in static, one-time period setup, and the authors coined the term *Acceptability Index* (AI) as a mathematical terminology for MOPs. Our goal is to elevate the mathematical framework for studying AIs to dynamical, multiperiod setup, where cash flows are considered as random processes, and one needs to assess their acceptability consistently in time. In particular, we are concerned not just with the total cumulative terminal value of the cash flow as seen from the initial time of the investment process, but also with all remaining cumulative cash flows between each intermediate time and the terminal time of the investment process.

Thus, in a sense, our program is analogous to the one of those researchers (cf. Cvitanic and Karatzas 1999; Cheridito, Delbaen, and Kupper 2004; Cheridito, Delbaen, and Kupper 2005, 2006; Frittelli and Gianin 2004; Riedel 2004; Detlefsen and Scandolo 2005; Roorda, Schumacher, and Engwerda 2005; Frittelli and Scandolo 2006; Weber 2006; Artzner et al. 2007; Jobert and Rogers 2008; Tutsch 2008; Bion-Nadal 2009; Delbaen, Peng, and Gianin 2010) who are studying dynamic risk measures. Moreover, as will be seen later in Section 4, there is a duality relationship between dynamic (coherent) acceptability indices (DCAIs) and dynamic (coherent) risk measures (DCRMs).

The paper is organized as follows. In Section 2, we summarize the main results of Cherny and Madan (2009). This is done for the convenience of the reader, but also to give the flavor of the duality between acceptability indices and risk measures, that will be generalized to the dynamic framework in the subsequent sections. In Section 3, we present the definition of a DCAI. We devote some time to discussion of the properties of DCAI from the definition, putting special emphasis on discussion of various forms of the dynamic consistency property.

Section 4 first introduces the concept of the DCRM, specific for our needs, and then proceeds to study the duality between families of such measures and DCAI. In the process, we discuss the dynamic consistency property of a DCRM, and we relate our findings to the results from existing literature.

In Section 5, we provide characterization of a DCRM in terms of a so-called dynamically consistent sequence of sets of probability measures, thereby providing an additional perspective at DCAIs.

Section 6 is dedicated to discussion of some abstract examples of dynamic MOPs, as well as some specific examples of dynamic MOPs derived form the classical ones, such as GLR and RAROC. In particular, we show that the dynamic version of GLR is a DCAI, whereas the dynamic version of RAROC is not.

2. STATIC ACCEPTABILITY INDICES

In this section, we will briefly review the theory of static acceptability indices developed by Cherny and Madan (2009).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and denote by $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ the space of all bounded random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. The random variable $X \in L^\infty$ can be regarded

[1] There exists a vast literature that studies measures of risk in a general mathematical framework.

as discounted *terminal* cash flow of a zero-cost self-financed portfolio. By definition, an AI is a map $\alpha\colon L^\infty \to [0, +\infty]$. The value $\alpha(X)$ should be understood as the degree of acceptability of a cash flow X; in a sense, it represents a measure of efficiency of the cash flow. A larger index indicates better performance, with $\alpha(X) = +\infty$ for X being an "arbitrage opportunity"; in particular, if the cash flow is strictly positive, then $\alpha(X) = +\infty$.

AI as such is a too broad concept, and it may not fulfill certain practically desirable properties. That is why Cherny and Madan (2009) focused their attention on a more specific concept of the CAI.

DEFINITION 2.1. An AI is called coherent if the following properties are satisfied:

(S1) *Monotonicity*. If $X \le Y$, then $\alpha(X) \le \alpha(Y)$;
(S2) *Scale invariance*. For every $X \in L^\infty$ and $\lambda > 0, \alpha(\lambda X) = \alpha(X)$;
(S3) *Quasi-concavity*. If $\alpha(X) \ge x, \alpha(Y) \ge x$, then $\alpha(\lambda X + (1 - \lambda)Y) \ge x$ for all $\lambda \in [0, 1]$;
(S4) *Fatou property*. If $|X_n| \le 1, \alpha(X_n) \ge x$ for all $n \ge 1$, and $X_n \to X$, as $n \to \infty$, in probability, then $\alpha(X) \ge x$.

The above properties have natural financial interpretation. For example, (S1) states that if Y dominates $X - \mathbb{P}$ almost surely,[2] then Y is acceptable at least at the same level as X is; (S2) implies that cash flows with the same direction of trade have the same level of acceptance. Quasi-concavity (S3) implies that a diversified portfolio performs at higher level than its components; to see this, it is enough to take $x = \min\{\alpha(X), \alpha(Y)\}$. Fatou property (S4) is a technical continuity property, which is used for constructing the duality between CAIs and CRMs.

It can be shown that SR, defined as $\mathrm{SR}(X) := \frac{\mathrm{E}(X)-r}{\mathrm{STD}(X)}$, where $\mathrm{STD}(X)$ is the standard deviation of X and r is the (constant) interest rate, does not satisfy the monotonicity property (S1), and hence it is not a CAI. The GLR, given by $\mathrm{GLR}(X) := \mathbb{E}(X)/\mathbb{E}(\max\{-X, 0\})$ if $X > 0$, and zero otherwise, is a CAI. Other MOPs such as RAROC, AIT, AIW, AIMIN, AIMAX, AIMINMAX, AIMAXMIN, etc., have been also studied by Cherny and Madan (2009). Moreover, the authors proved the following representation theorem.

THEOREM 2.2. *A map $\alpha\colon L^\infty \to [0, +\infty]$, unbounded from above, is a CAI if and only if there exists a family $(\mathcal{D}_x)_{x\in[0,+\infty]}$ of sets of probability measures, such that $\mathcal{D}_x \subset \mathcal{D}_y$ for $x \le y$, and α admits the following representation*

$$\alpha(X) = \sup\left\{x \in [0, +\infty) \ : \ \inf_{Q\in\mathcal{D}_x} \mathbb{E}_Q[X] \ge 0\right\}, \tag{2.1}$$

where $\inf \emptyset = \infty$ *and* $\sup \emptyset = 0$.

Thus, any CAI can be characterized by an increasing family of sets of probability measures. This family of probability measures can be seen as generalized scenarios as described by Artzner et al. (1999), or a set of supporting kernels as discussed by Cherny and Madan (2009). Moreover, there is a strong relationship between CAI and CRM, a concept introduced by Artzner et al. (1997, 1999).

[2]In this we make a standing assumption that Ω is finite and that $\mathbb{P}$ is strictly positive. Thus, our statements regarding relations between random variables will hold point-wise. In particular, $Y \ge X$ will mean that Y dominates X for every $\omega \in \Omega$.

DEFINITION 2.3. A function $\rho : L^\infty \to \mathbb{R}$ is called CRM if the following properties are satisfied:

(R1) *Monotonicity.* If $X \leq Y$, then $\rho(X) \geq \rho(Y)$;
(R2) *Positive homogeneity.* $\rho(\lambda X) = \lambda\rho(X)$, for every $X \in L^\infty$ and $\lambda \geq 0$;
(R3) *Translation property.* $\rho(X + k) = \rho(X) - k$, for every $X \in L^\infty$ and $k \in \mathbb{R}$;
(R4) *Subadditivity.* $\rho(X + Y) \leq \rho(X) + \rho(Y)$, for every $X, Y \in L^\infty$.

Traditional Value at Risk $\mathrm{VaR}_\alpha(X) := \inf\{m \in \mathbb{R} \mid \mathbb{P}[X+m < 0] \leq \alpha\}$, while very popular, is not a CRM since it lacks the subadditivity property (R4), which corresponds to the diversification property in finance. In contrast, the Tail Value at Risk (also called Tail Conditional Expectation), defined as $\mathrm{TVaR}_\alpha(X) := -\inf_{\mathbb{Q}\in\mathcal{Q}_\alpha} \mathbb{E}_{\mathbb{Q}}[X]$, where $\alpha \in (0, 1]$ and $\mathcal{Q}_\alpha$ is the set of probability measures absolutely continuous with respect to $\mathbb{P}$ such that $d\mathbb{Q}/d\mathbb{P} \leq \alpha^{-1}$, is a CRM. So is the Weighted Value at Risk, $WVaR_\mu(X) := \int_{(0,1]} TVaR_\alpha(X)\mu(d\alpha)$, where μ is a probability measure on $(0, 1]$. The following representation theorem is established in Artzner et al. (1999) for finite Ω, and generalized to a general probability space by Delbaen (2002) and Carr, Geman, and Madan (2001).

THEOREM 2.4. *A function $\rho : L^\infty \to \mathbb{R}$ is a CRM if and only if*

$$\rho(X) = \sup\{\mathbb{E}_{\mathbb{Q}}[-X] \ : \ \mathbb{Q} \in \mathcal{P}\} \tag{2.2}$$

for a certain set $\mathcal{D}$ of probability measures absolutely continuous with respect to $\mathbb{P}$.

Note that by (2.1) and the above theorem, every CAI can be characterized in terms of an increasing family of CRMs.

The theory of static risk measures has been explored and extended by many researchers; to mention just a few of them: Föllmer and Schied (2002a,b) and Frittelli and Gianin (2002) generalized the concept of CRMs to convex and monetary risk measures; Cheridito and Li (2009) studied generalized measures on Orlicz Hearts; law-invariant risk measures have been investigated by Kusuoka (2001) and Frittelli and Gianin (2005); for a systematic discussion on static risk measures we refer reader to the monographs by Delbaen (2000) and Föllmer and Schied (2004, chapter 4).

3. DYNAMIC COHERENT ACCEPTABILITY INDICES

As has been already stated, the dynamic acceptability indices are meant to assess performance of a cash-flow accounting for newly acquired information when time progresses. Of course, one may attempt to use for this purpose a sequence of static (one-period) acceptability indices. However, by doing this one may end up with a sequence of measurements that are not consistent in time, in the sense to be explained below (cf. property D7). The motivation for developing a theory of DCAIs, as presented in this paper, was to create a systematic mathematical framework to provide performance measurements consistently in time.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a finite underlying probability space, and let $\mathcal{T} = \{0, 1, 2, \ldots, T\}$ be a finite set of time instants. We assume that $\mathbb{P}$ is of full support. We endow the underlying probability space with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^T$. For each $t \in \mathcal{T}$ and $\mathcal{F}_t \in \mathbb{F}$, there exists a partition of Ω, say $\{P_1^t, P_2^t, \ldots, P_{n_t}^t\}$, that generates $\mathcal{F}_t$.

A cash flow, also called dividend process, denoted as $D = \{D_t(\omega)\}_{t=0}^T$, is any real valued random process adapted to the filtration $\mathbb{F}$. We denote by $\mathcal{D}$ the set of all cash flows. In addition, we denote by $\mathcal{P}$ the set of all probability measures that are absolutely continuous with respect to $\mathbb{P}$, and by $\mathcal{P}^e$ the set of all probability measures equivalent to $\mathbb{P}$. Also, c will denote a generic constant, and m will denote a generic random variable. Finally, a standing (financial type) assumption, which we make without loss of generality, is that the interest rates are zero.

DEFINITION 3.1. A DCAI is a function $\alpha : \mathcal{T} \times \mathcal{D} \times \Omega \to [0, +\infty]$ that satisfies the following properties:

(D1) *Adaptiveness.* For any $t \in \mathcal{T}$ and $D \in \mathcal{D}$, $\alpha_t(D)$ is $\mathcal{F}_t$-measurable;
(D2) *Independence of the past.* For any $t \in \mathcal{T}$ and $D, D' \in \mathcal{D}$, if there exists $A \in \mathcal{F}_t$ such that $1_A D_s = 1_A D'_s$ for all $s \geq t$, then $1_A \alpha_t(D) = 1_A \alpha_t(D')$;
(D3) *Monotonicity.* For any $t \in \mathcal{T}$ and $D, D' \in \mathcal{D}$, if $D_s(\omega) \geq D'_s(\omega)$ for all $s \geq t$ and $\omega \in \Omega$, then $\alpha_t(D, \omega) \geq \alpha_t(D', \omega)$ for all $\omega \in \Omega$;
(D4) *Scale invariance.* $\alpha_t(\lambda D, \omega) = \alpha_t(D, \omega)$ for all $\lambda > 0$, $D \in \mathcal{D}$, $t \in \mathcal{T}$, and $\omega \in \Omega$;
(D5) *Quasi-concavity.* If $\alpha_t(D, \omega) \geq x$ and $\alpha_t(D', \omega) \geq x$ for some $t \in \mathcal{T}$, $\omega \in \Omega$, $D, D' \in \mathcal{D}$, and $x \in (0, +\infty]$, then $\alpha_t(\lambda D + (1 - \lambda)D', \omega) \geq x$ for all $\lambda \in [0, 1]$;
(D6) *Translation invariance.* $\alpha_t(D + m1_{\{t\}}, \omega) = \alpha_t(D + m1_{\{s\}}, \omega)$ for every $t \in \mathcal{T}$, $D \in \mathcal{D}$, $\omega \in \Omega$, $s \geq t$ and every $\mathcal{F}_t$-measurable random variable m;
(D7) *Dynamic consistency.* For any $t \in [0, \ldots, T - 1]$ and $D, D' \in \mathcal{D}$, if $D_t(\omega) \geq 0 \geq D'_t(\omega)$ for all $\omega \in \Omega$, and there exists a nonnegative $\mathcal{F}_t$-measurable random variable m such that $\alpha_{t+1}(D, \omega) \geq m(\omega) \geq \alpha_{t+1}(D', \omega)$ for all $\omega \in \Omega$, then $\alpha_t(D, \omega) \geq m(\omega) \geq \alpha_t(D', \omega)$ for all $\omega \in \Omega$.

Property (D1) is a natural property in a dynamic setup and it assumes that a DCAI is adapted to the same information flow $\{\mathcal{F}_t\}_{t\geq 0}$ as is any cash flow $D \in \mathcal{D}$.

(D2) postulates that in the dynamic context the current measurement of performance of a cash flow D only accounts for future payoffs. To decide, at any given point of time, whether one should hold on to a position generating the cash flow D, one may want to compare the measurement of the performance of the future payoffs (provided by DCAI at this point of time) to already known past payoffs.

Properties (D3)–(D5) are naturally inherited from the static case (cf. Section 2).

Translation invariance (D6) implies that if a known dividend m is added to D at time t (today), or at any future time $s \geq t$, then all such adjusted cash flows are accepted today at the same level.

Dynamic consistency (D7) is the property in the dynamic setup which relates the values of the index between two consecutive days in a consistent manner. It can be interpreted from a financial point of view as follows: if a portfolio has a nonnegative cash flow today, then we accept this portfolio today at least at the same level as we would accept it tomorrow; similarly, if the today's cash flow is nonpositive the acceptance level today can not be larger than the level of acceptance tomorrow.

For technical reasons, which will become clear later, we assume that for every DCAI α, and for every $t \in \mathcal{T}$ and $\omega \in \Omega$, there exist two portfolios $D, D' \in \mathcal{D}$ such that $\alpha_t(D, \omega) = +\infty$ and $\alpha_t(D', \omega) = 0$. We say that DAI α is *normalized*.

Note that normalization will exclude the degenerate examples of acceptability indices such as a constant index over all states, times, and portfolios. Moreover, one can show that a normalized index gets value infinity for every strictly positive cash flow and value zero if the cash flow is strictly negative:

$$\alpha_t(D^{c,s}) = +\infty \quad \text{for } c > 0, \quad \text{and} \quad \alpha_t(D^{c,s}) = 0 \ \text{ for } c < 0, \quad \text{for all} \quad t \in \mathcal{T},$$

where for any $\omega \in \Omega$ and $s \geq t$, $D^{c,s}(r, \omega) = c$ for $r = s$ and zero otherwise.

For normalized DCAI we have equivalent forms of property (D7). In fact, one can show that under normalization, the set of properties (D1)–(D7) is equivalent to either the set (D1)–(D7-I) or the set (D1)–(D7-II), where

(D7-I) For a given $t \geq 0$ and $D, D' \in \mathcal{D}$, if $D_t(\omega) = D'_t(\omega) = 0$ for all $\omega \in \Omega$, and there exists a nonnegative $\mathcal{F}_t$-measurable random variable m such that $\alpha_{t+1}(D, \omega) \geq m(\omega) \geq \alpha_{t+1}(D', \omega)$ for all $\omega \in \Omega$, then $\alpha_t(D, \omega) \geq m(\omega) \geq \alpha_t(D', \omega)$ for all $\omega \in \Omega$.
(D7-II) For a given $t \geq 0$ and $D \in \mathcal{D}$, if $D_t(\omega) = 0$ for all $\omega \in \Omega$, then

$$1_A \min_{\omega \in A} \alpha_{t+1}(D, \omega) \leq 1_A \alpha_t(D) \leq 1_A \max_{\omega \in A} \alpha_{t+1}(D, \omega),$$

for all $A \in \mathcal{F}_t$.

Finally we want to mention that (D3) and (D7) can be equivalently replaced in the definition of DCAI by the following two properties:

(D3-I) For $D, D' \in \mathcal{D}$, if there exists $A \in \mathcal{F}_t$ such that $1_A D_s \geq 1_A D'_s$ for all $s \geq t$, then $1_A \alpha_t(D) \geq 1_A \alpha_t(D')$;
(D7-III) For $D, D' \in \mathcal{D}$, if there exist $A \in \mathcal{F}_t$ and a nonnegative $\mathcal{F}_t$-measurable random variable m, such that $1_A D_t \geq 0 \geq 1_A D'_t$ and $1_A \alpha_{t+1}(D) \geq 1_A m \geq 1_A \alpha_{t+1}(D')$, then $1_A \alpha_t(D) \geq 1_A m \geq 1_A \alpha_t(D')$.

4. CHARACTERIZATION OF DYNAMIC CAI BY A FAMILY OF DYNAMIC CRMs

As mentioned in Section 2, there is a strong relationship between CAIs and CRMs. In fact, as seen from Theorems 2.2 and 2.4, any CAI α can be represented in terms of a family of CRMs ρ^x, $x \geq 0$:

$$\alpha(D) = \sup\{x \in [0, +\infty) : \rho^x(D) \leq 0\}\,. \tag{4.1}$$

Looking at (4.1) one might think that a natural approach to constructing a DCAI would be to use this representation but to replace the static CRMs in (4.1) by their dynamic counterpart. The representation (4.8) that we derive below shows that this is indeed the case. The delicate issue however is, what family of DCRMs should be used. It turns out that to produce a DCAI satisfying a financially acceptable set of dynamic properties, one needs to use a carefully crafted family of DCRMs. In this section we introduce such a family of DCRMs and we compare our definition of DCRMs with analogous ones that have been already studied in the literature.

4.1. Definition of DCRM

DEFINITION 4.1. DCRM is a function $\rho : \{0, \ldots, T\} \times \mathcal{D} \times \Omega \to \mathbb{R}$ that satisfies the following properties:

(A1) *Adaptiveness.* $\rho_t(D)$ *is* $\mathcal{F}_t$-measurable for all $t \in \mathcal{T}$ and $D \in \mathcal{D}$;
(A2) *Independence of the past.* If $1_A D_s = 1_A D'_s$ for some $t \in \mathcal{T}$, $D, D' \in \mathcal{D}$, and $A \in \mathcal{F}_t$ and for all $s \geq t$, then $1_A \rho_t(D) = 1_A \rho_t(D')$;

(A3) *Monotonicity.* If $D_s(\omega) \geq D'_s(\omega)$ for some $t \in \mathcal{T}$ and $D, D' \in \mathcal{D}$, and for all $s \geq t$ and $\omega \in \Omega$, then $\rho_t(D, \omega) \leq \rho_t(D', \omega)$ for all $\omega \in \Omega$;
(A4) *Homogeneity.* $\rho_t(\lambda D, \omega) = \lambda \rho_t(D, \omega)$ for all $\lambda > 0$, $D \in \mathcal{D}$, $t \in \mathcal{T}$, and $\omega \in \Omega$;
(A5) *Subadditivity.* $\rho_t(D + D', \omega) \leq \rho_t(D, \omega) + \rho_t(D', \omega)$ for all $t \in \mathcal{T}$, $D, D' \in \mathcal{D}$, and $\omega \in \Omega$;
(A6) *Translation invariance.* $\rho_t(D + m1_{\{s\}}) = \rho_t(D) - m$ for every $t \in \mathcal{T}$, $D \in \mathcal{D}$, $\mathcal{F}_t$-measurable random variable m, and all $s \geq t$;
(A7) *Dynamic consistency.* $1_A\big(\min_{\omega \in A} \rho_{t+1}(D, \omega) - D_t\big) \leq 1_A \rho_t(D) \leq 1_A\big(\max_{\omega \in A} \rho_{t+1}(D, \omega) - D_t\big)$, for every $t \in \{0, 1, \ldots, T-1\}$, $D \in \mathcal{D}$ and $A \in \mathcal{F}_t$.

We want to mention that our definition of DCRM differs from the definition given in previous studies essentially only by the dynamic consistency property. For sake of completeness, we will present here how property (A7) relates to other forms of dynamic consistency of risk measures (for processes).

(A7-I) If $D_t = D'_t$, and $\rho_{t+1}(D) = \rho_{t+1}(D')$ for some $t \in \{0, 1, \ldots, T-1\}$, and $D, D' \in \mathcal{D}$, then $\rho_t(D) = \rho_t(D')$;
(A7-II) $\rho_t(D) = \rho_t(-\rho_{t+1}(D)1_{\{t+1\}}) - D_t$ for all times $t = 0, 1, \ldots, T-1$ and positions $D \in \mathcal{D}$.
(A7-III) $\rho_t(D) \leq \rho_t(-\rho_{t+1}(D)1_{t+1}) - D_t$ for all $D \in \mathcal{D}$, $t \in \{0, 1, \ldots, T-1\}$,
(A7-IV) $\rho_t(D) \geq \rho_t(-\rho_{t+1}(D)1_{t+1}) - D_t$ for all $D \in \mathcal{D}$, $t \in \{0, 1, \ldots, T-1\}$,
(A7-V) if $D_t = 0$, and $\rho_{t+1}(D) \leq 0$ for some $t \in \{0, 1, \ldots, \}$ and $D \in \mathcal{D}$, then $\rho_t(D) \leq 0$.

Property (A7-I) is the dynamic consistency property for DCRM defined by Riedel (2004). Property (A7-II) is the version of the dynamic programming principle (also called recursiveness), introduced by Cheridito et al. (2006), adapted to the setup of our paper, that is, it is stated in terms of dividend processes rather than value process as in Cheridito et al. (2006). Properties (A7-I) and (A7-II) are equivalent, and they are also sometimes called *strong dynamic consistency property*. To the best of our knowledge, properties (A7-III) and (A7-IV) were first introduced in the context of random processes by Acciaio, Föllmer, and Penner (2010), and they were called *acceptance* and *rejection consistency*, respectively. In the same paper, Acciaio et al. introduced condition (A7-V) and they called it *weakly acceptance consistent*.

For corresponding definitions in case of random variables rather than random processes we refer to the survey paper by Acciaio and Penner (2011) and references therein.

It is easy to show that the dynamic consistency condition (A7) is stronger than (A7-V), and it is weaker than (A7-I) or (A7-II). Also note that since conditions (A7-II) and (A7-III) taken together are equivalent to (A7-II), then, taken together they imply (A7). However, the inverse implication is not necessarily true.

We conclude this subsection with the following result.

PROPOSITION 4.2. *If ρ is a DCRM, then $\rho_t(c1_{\{s\}}, \omega) = -c$, for all $c \in \mathbb{R}$, $t \in \mathcal{T}$, $\omega \in \Omega$ and $s \geq t$.*

Proof. Given some fixed $t \in \mathcal{T}$ and $\omega \in \Omega$, denote by $\lambda := \rho_t(0, \omega)$. Then, by translation invariance (A6) of ρ, we deduce

$$\rho_t(c1_{\{s\}}, \omega) = \rho_t(0, \omega) - c = \lambda - c, \tag{4.2}$$

for all $c \in \mathbb{R}$. In particular, for $c = 1$, we have $\rho_t(1_{\{s\}}, \omega) = \lambda - 1$. Hence, by (A4)-homogeneity of ρ, it follows that $\rho_t(c_u 1_{\{s\}}, \omega) = c_u \rho_t(1_{\{s\}}, \omega) = c_u(\lambda - 1)$, for all $c_u > 0$. Combining this with (4.2) we get $\lambda - c_u = c_u\lambda - c_u$, and consequently $\lambda(1 - c_u) = 0$, for arbitrary positive c_u. Thus, we conclude that $\lambda = 0$, and hence $\rho_t(0, \omega) = 0$. With this, by (4.2), the proposition follows. □

Note that, in particular, $\rho_t(0) = 0$, for all $t \in \mathcal{T}$.

4.2. Duality between DCAI and DCRM

We start this section with several definitions that will be used in the main results derived here.

DEFINITION 4.3. A family of DCRMs $(\rho^x)_{x\in(0,+\infty)}$ is called increasing if $\rho_t^x(D, \omega) \geq \rho_t^y(D, \omega)$, for all $x \geq y > 0$, $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$.

DEFINITION 4.4. A dynamic AI α is called right-continuous if $\lim_{c\to 0^+} \alpha_t(D + c1_{\{t\}}, \omega) = \alpha_t(D, \omega)$, for all $t \in \mathcal{T}$, $D \in \mathcal{D}$, and $\omega \in \Omega$.

DEFINITION 4.5. A family of DCRMs $(\rho^x)_{x\in(0,+\infty)}$ is called left-continuous at $x_0 > 0$, if $\lim_{x\to x_0^-} \rho_t^x(D, \omega) = \rho_t^{x_0}(D, \omega)$, for all $t \in \mathcal{T}$, $D \in \mathcal{D}$, and $\omega \in \Omega$.

THEOREM 4.6. *Assume that α is a normalized DCAI. Then, the set of functions ρ^x, $x \in \mathbb{R}$, defined by*

$$\rho_t^x(D, \omega) := \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}, \tag{4.3}$$

for all $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$, is an increasing, left-continuous family of DCRMs.

Proof. First we will show that ρ^x defined by (4.3) is well defined. Since α is normalized, for all $t \in \mathcal{T}$, $D \in \mathcal{D}$, there exist two finite constants $c_u^{t,D}$ and $c_l^{t,D}$ such that

$$\alpha_t\big(D + c_u^{t,D}1_{\{t\}}, \omega\big) = +\infty \quad \text{and} \quad \alpha_t\big(D + c_l^{t,D}1_{\{t\}}, \omega\big) = 0,$$

for all $\omega \in \Omega$. Hence, for every $x \in (0, +\infty)$, the set $\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}$ is not empty, and $c_l^{t,D} \leq \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}$. From here we conclude that infimum from (4.3) is finite, and hence ρ^x is well defined.

Next we will show that ρ^x, $x \in (0, +\infty)$, satisfies properties (A1)–(A7). By (D1)-adaptiveness and (D2)-independence of the past of α, property (A1) and (A2) for ρ^x, $x \in \mathbb{R}$, follow immediately.

Assume that $t \in \mathcal{T}$ and $D, D' \in \mathcal{D}$ are such that $D_s(\omega) \geq D'_s(\omega)$ for all $s \geq t$ and $\omega \in \Omega$. Then $(D + c1_{\{t\}})_s(\omega) \geq (D' + c1_{\{t\}})_s(\omega)$ for $s \geq t, \omega \in \Omega$, and $c \in \mathbb{R}$, and by (D3), monotonicity of α

$$\alpha_t(D + c1_{\{t\}}, \omega) \geq \alpha_t(D' + c1_{\{t\}}, \omega), \tag{4.4}$$

for all $c \in \mathbb{R}$ and $\omega \in \Omega$. From here, we deduce the following inclusion

$$\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\} \supseteq \{c \in \mathbb{R} : \alpha_t(D' + c1_{\{t\}}, \omega) \geq x\}.$$

Taking infimum of both sets, $(A3)$ follows. Similarly, the homogeneity of ρ^x follows from the scale invariance of α.

Next we show that ρ^x satisfies $(A5)$. Let $t \in \mathcal{T}$, $D, D' \in \mathcal{D}$ and $\omega \in \Omega$, and let us take $c_1, c_2 \in \mathbb{R}$ such that

$$\alpha_t(D + c_1 1_{\{t\}}, \omega) \geq x, \quad \alpha_t(D' + c_2 1_{\{t\}}, \omega) \geq x.$$

Then, by (D5), quasi-concavity of α, we have

$$\alpha_t\left(\frac{1}{2}D + \frac{1}{2}c_1 1_{\{t\}} + \frac{1}{2}D' + \frac{1}{2}c_2 1_{\{t\}}, \omega\right) \geq x,$$

and therefore by (D4), scale invariance of α, we get $\alpha_t(D + D' + (c_1 + c_2)1_{\{t\}}, \omega) \geq x$. This implies that $c_1 + c_2 \in \{c \in \mathbb{R} : \alpha_t(D + D' + c1_{\{t\}}, \omega) \geq x\}$. Hence,

$$\begin{aligned} c_1 + c_2 &\geq \inf\{c \in \mathbb{R} : \alpha_t(D + D' + c1_{\{t\}}, \omega) \geq x\} \\ &= \rho_t^x(D + D', \omega). \end{aligned} \tag{4.5}$$

Note that the above inequality holds true for all $c_1 \in \{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}$ and $c_2 \in \{c \in \mathbb{R} : \alpha_t(D' + c1_{\{t\}}, \omega) \geq x\}$. By taking infimum in (4.5), first with respect to c_1, and then with respect to c_2, we have, $\rho_t^x(D, \omega) + \rho_t^x(D', \omega) \geq \rho_t^x(D + D', \omega)$, and hence $(A5)$ is checked.

Now we will show that ρ^x satisfies (A6), translation invariance. Fix an $\omega^0 \in \Omega$, $t \in \mathcal{T}$, $D \in \mathcal{D}$ and an $\mathcal{F}_t$-measurable random variable m. Denote by P_i^t the unique element of partition of $\mathcal{F}_t$ such that $\omega^0 \in P_i^t$. This yields that the cash flows $m1_{\{t\}}$ and $m(\omega^0)1_{\{t\}}$ agree on the set P_i^t, and for all times $s \geq t$. Then, for any constant $c \in \mathbb{R}$, we have

$$1_{P_i^t}(D + m1_t + c1_{\{t\}})_s = 1_{P_i^t}(D + m(\omega^0)1_{\{t\}} + c1_{\{t\}})_s, \quad \text{for} \quad s \geq t.$$

By (D2), independence of the past of α, we have

$$1_{P_i^t}\alpha_t(D + m1_t + c1_{\{t\}}) = 1_{P_i^t}\alpha_t(D + m(\omega^0)1_{\{t\}} + c1_{\{t\}}).$$

Since m is $\mathcal{F}_t$-measurable, by (D6), translation invariance of α, we have

$$\alpha_t(D + m1_s + c1_{\{t\}}, \omega^0) = \alpha_t(D + m1_t + c1_{\{t\}}, \omega^0), \quad \text{for all} \quad s \geq t.$$

Combining the above with (4.3), we deduce

$$\begin{aligned} \rho_t^x(D + m1_{\{s\}}, \omega^0) &= \inf\{c \in \mathbb{R} : \alpha_t(D + m1_{\{s\}} + c1_{\{t\}}, \omega^0) \geq x\} \\ &= \inf\{c \in \mathbb{R} : \alpha_t(D + m1_{\{t\}} + c1_{\{t\}}, \omega^0) \geq x\} \\ &= \inf\{c \in \mathbb{R} : \alpha_t(D + m(\omega^0)1_{\{t\}} + c1_{\{t\}}, \omega^0) \geq x\} \\ &= \inf\{m(\omega^0) + c \in \mathbb{R} : \alpha_t(D + (m(\omega^0) + c)1_{\{t\}}, \omega^0) \geq x\} - m(\omega^0) \\ &= \rho_t^x(D, \omega) - m(\omega^0). \end{aligned}$$

Since ω^0 is arbitrarily chosen in Ω, we obtain $\rho_t^x(D+m1_{\{s\}}) = \rho_t^x(D) - m$, for all $s \geq t$, and (A6) is checked.

Next we will show that ρ^x satisfies ($A7$), dynamic consistency. Assume that $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $A \in \mathcal{F}_t$ are fixed, and denote by $c_{\min}^{t,D,A} := \min_{\omega \in A} \rho_{t+1}^x(D, \omega)$ and $c_{\max}^{t,D,A} := \max_{\omega \in A} \rho_{t+1}^x(D, \omega)$. Then $\alpha_{t+1}(D + c_0 1_{\{t+1\}}, \omega) < x$, for all $\omega \in A$ and for any $c_0 < c_{\min}^{t,D,A}$. Due to the finiteness of the probability space Ω, there exists a number $\epsilon_A > 0$, such that $\alpha_{t+1}(D + c_0 1_{\{t+1\}}, \omega) \leq x - \epsilon_A$, for all $\omega \in A$. By (D2), independent of the past of α, we have

$$\alpha_{t+1}(D - D_t 1_{\{t\}} + c_0 1_{\{t+1\}}, \omega) = \alpha_{t+1}(D + c_0 1_{\{t+1\}}, \omega) \leq x - \epsilon_A,$$

for all $\omega \in A$. Since, $1_A(D - D_t 1_{\{t\}} + c_0 1_{\{t+1\}})_t = 1_A(D_t - D_t) = 0$, then, by (D7)

$$\alpha_t(D - D_t 1_{\{t\}} + c_0 1_{\{t+1\}}, \omega) \leq x - \epsilon_A, \quad \omega \in A.$$

Consequently, since c_0 is a constant, by (D6)

$$\begin{aligned}\alpha_t(D + (c_0 - D_t)1_{\{t\}}, \omega) &= \alpha_t(D - D_t 1_{\{t\}} + c_0 1_{\{t\}}, \omega) \\ &= \alpha_t(D - D_t 1_{\{t\}} + c_0 1_{\{t+1\}}, \omega) \\ &\leq x - \epsilon_A < x,\end{aligned}$$

for all $\omega \in A$ and $c_0 < c_{\min}^{t,D,A}$. By the definition of ρ^x, we get

$$\rho_t^x(D, \omega) = \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\} \geq c_0 - D_t(\omega),$$

for all $\omega \in A$ and $c_0 < c_{\min}^{t,D,A}$. Hence, $\rho_t^x(D, \omega) \geq c_{\min}^{t,D,A} - D_t(\omega)$, or equivalently $1_A \rho_t^x(D) \geq 1_A(\min_{\omega \in A} \rho_{t+1}^x(D, \omega) - D_t)$. Similarly, one can show that

$$1_A \rho_t^x(D) \leq 1_A\big(\max_{\omega \in A} \rho_{t+1}(D, \omega) - D_t\big),$$

and thus (A7) is established.

All the above imply that ρ^x is a DCRM for every $x > 0$.

Monotonicity of ρ^x with respect to x follows immediately from (4.3) and the inclusion

$$\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\} \subseteq \{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq y\}, \quad x \geq y > 0.$$

Finally, we will show that $(\rho^x)_{x \in (0,+\infty)}$ is left-continuous. Let x_0 be any positive number. Then,

$$\inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x_0\} \geq \lim_{x \to x_0^-} \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}. \tag{4.6}$$

If the above inequality holds strictly, then there exists a constant c_0 such that,

$$\inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x_0\} > c_0 > \lim_{x \to x_0^-} \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}. \tag{4.7}$$

Note that $\inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}$ is a nondecreasing function with respect to x. Therefore, the second inequality in (4.7) implies that $c_0 > \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}}, \omega) \geq x\}$, for all $x < x_0$. Hence, by (D3), monotonicity of α, $\alpha_t(D + c_0 1_{\{t\}}, \omega) \geq x$, for all $x < x_0$, and thus $\alpha_t(D + c_0 1_{\{t\}}, \omega) \geq \lim_{x \to x_0^-} x = x_0$. On the other hand, by the first inequality in (4.7), we deduce that, $\alpha_t(D + c_0 1_{\{t\}}, \omega) < x_0$. Contradiction. Therefore, we should have strict equality in (4.6).

This completes the proof. □

The next theorem shows the representation of a DCAI in terms of a family of DCRMs.

THEOREM 4.7. *Assume that $(\rho^x)_{x\in(0,+\infty)}$ is an increasing family of DCRMs. Then the function α defined as follows,*

$$\alpha_t(D, \omega) := \sup\left\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le 0\right\}, \tag{4.8}$$

for $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$, is a normalized, right-continuous, DCAI. Here, we assume $\sup \emptyset = 0$.

Proof. Note that the assumption $\sup \emptyset = 0$ guarantees that α from (4.8) is well defined and takes values in $[0, +\infty]$.

In the following, we will prove that α defined in (4.8) satisfies properties (D1)–(D7).

(D1)—adaptiveness, (D2)—independence of the past, (D4)—scale invariance, and (D6)—translation invariance follow immediately from the definition of α, and from adaptiveness (A1), independence of the past (A2), homogeneity (A4), and translation invariance (A6) of ρ^x, respectively.

Let $t \in \mathcal{T}$, $D, D' \in \mathcal{D}$, and assume that $D_s(\omega) \ge D'_s(\omega)$ for all $s \ge t$, and $\omega \in \Omega$. By (A3), monotonicity of ρ^x, we have

$$\rho_t^x(D) \le \rho_t^x(D'), \qquad \text{for all} \quad x > 0\,. \tag{4.9}$$

Note that, for any $x_0 \in \{x \in (0, +\infty) : \rho_t^x(D', \omega) \le 0\}$, we have $\rho_t^{x_0}(D', \omega) \le 0$, which combined with (4.9) implies $\rho_t^{x_0}(D, \omega) \le \rho_t^{x_0}(D', \omega) \le 0, \quad \omega \in \Omega$. Therefore,

$$\left\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le 0\right\} \supseteq \left\{x \in (0, +\infty) : \rho_t^x(D', \omega) \le 0\right\}$$

By taking supremum of both sides, (D3) follows.

Next we will prove that α is quasi-concave. For given $t \in \mathcal{T}$, and $x^0 \in (0, +\infty]$, if $D, D' \in \mathcal{D}$ are such that $\alpha_t(D, \omega) \ge x^0$, $\alpha_t(D', \omega) \ge x^0$, then, by definition (4.8) of α, and monotonicity of ρ^x in x, we conclude that for any $x < x^0$,

$$\rho_t^x(D, \omega) \le 0, \quad \rho_t^x(D', \omega) \le 0\,.$$

By (A4), homogeneity of ρ^x, we note that for any $\lambda \in [0, 1]$ and $x < x^0$,

$$\rho_t^x(\lambda D, \omega) = \lambda \rho_t^x(D, \omega) \le 0, \quad \rho_t^x((1-\lambda)D', \omega) = (1-\lambda)\rho_t^x(D', \omega) \le 0\,.$$

From here, by (A5), subadditivity of ρ^x, we get

$$\rho_t^x(\lambda D + (1-\lambda)D', \omega) \le \rho_t^x(\lambda D, \omega) + \rho_t^x((1-\lambda)D', \omega) \le 0,$$

for any $x < x^0$. Hence $\sup\{x \in (0, +\infty) : \rho_t^x(\lambda D + (1-\lambda)D', \omega) \le 0\} \ge x^0$, and thus, by definition (4.8) of α, we have, $\alpha(\lambda D + (1-\lambda)D', \omega) \ge x_0$. This yields quasi-concavity of α.

Assume that $D \in \mathcal{D}$, and m is an $\mathcal{F}_t$-measurable random variable. By (4.8) and (A6), we get

$$\begin{aligned}\alpha_t(D + m1_{\{s\}}, \omega) &= \sup\left\{x \in (0, +\infty) : \rho_t^x(D + m1_{\{s\}}, \omega) \leq 0\right\} \\ &= \sup\left\{x \in (0, +\infty) : \rho_t^x(D + m1_{\{t\}}, \omega) \leq 0\right\} \\ &= \alpha_t(D + m1_{\{t\}}, \omega),\end{aligned}$$

for all $s \geq t$ and $\omega \in \Omega$. Hence, α satisfies property (D6).

Now, let us show that α satisfies dynamic consistency property (D7). Assume that D, $D' \in \mathcal{D}$, and $t \in \mathcal{T}$ are such that $D_t(\omega) \geq 0 \geq D'_t(\omega)$ for all $\omega \in \Omega$, and there exists a nonnegative $\mathcal{F}_t$-measurable random variable m such that $\alpha_{t+1}(D, \omega) \geq m(\omega) \geq \alpha_{t+1}(D', \omega)$ for all $\omega \in \Omega$. By definition (4.8)

$$\sup\left\{x \in (0, +\infty) : \rho_{t+1}^x(D, \omega) \leq 0\right\} \geq m(\omega) \geq \sup\left\{x \in (0, +\infty) : \rho_{t+1}^x(D', \omega) \leq 0\right\},$$

for all $\omega \in \Omega$. Let us fix an $\bar{\omega} \in \Omega$, and denote by $\bar{c} := m(\bar{\omega})$. There exists a P_i^t such that $\bar{\omega} \in P_i^t$. From the above inequality, we conclude that for all $\omega \in P_i^t$

$$\sup\{x \in (0, +\infty) : \rho_{t+1}^x(D, \omega) \leq 0\} \geq \bar{c} \geq \sup\{x \in (0, +\infty) : \rho_{t+1}^x(D', \omega) \leq 0\}.$$

Then, for all $c' > \bar{c}$ and $\omega \in P_i^t$, $c' > \sup\{x \in (0, +\infty) : \rho_{t+1}^x(D', \omega) \leq 0\}$, which consequently implies that

$$\rho_{t+1}^{c'}(D', \omega) > 0. \tag{4.10}$$

Also note that $\sup\{x \in (0, +\infty) : \rho_{t+1}^x(D, \omega) \leq 0\} > c$, for any $c < \bar{c}$. By monotonicity of ρ^x with respect to x, we have $\rho_{t+1}^c(D, \omega) \leq 0,\ \omega \in P_i^t$. Due to the finiteness of Ω, (4.10) implies that $\min_{\omega \in P_i^t} \rho_{t+1}^{c'}(D', \omega) > 0$, for all $c' > \bar{c}$. Using (A7), dynamic consistency of ρ^x, we get the following

$$\begin{aligned}1_{P_i^t}\rho_t^{c'}(D') &\geq 1_{P_i^t}(\min_{\omega \in P_i^t} \rho_{t+1}^{c'}(D', \omega) - D'_t) \\ &= 1_{P_i^t} \min_{\omega \in P_i^t} \rho_{t+1}^{c'}(D', \omega) - 1_{P_i^t} D'_t, \quad c' > \bar{c}.\end{aligned}$$

Equivalently,

$$\rho_t^{c'}(D', \omega) = \min_{\omega \in P_i^t} \rho_{t+1}^{c'}(D', \omega) - D'_t(\omega) > -D'_t(\omega) \geq 0, \tag{4.11}$$

for all $\omega \in P_i^t$, and $c' > \bar{c}$.

If $\bar{c} < \sup\{x \in (0, +\infty) : \rho_t^x(D', \omega') \leq 0\}$, for some $\omega' \in P_i^t$, then there exists a constant c^0 such that

$$\bar{c} < c^0 < \sup\left\{x \in (0, +\infty) : \rho_t^x(D', \omega') \leq 0\right\}.$$

This implies that $\rho_t^{c^0}(D', \omega') \leq 0$, that contradicts with (4.11). Therefore,

$$\bar{c} \geq \sup\left\{x \in (0, +\infty) : \rho_t^x(D', \omega) \leq 0\right\},$$

and by (4.8), we have

$$\bar{c} \geq \alpha_t(D', \omega), \quad \omega \in P_i^t. \tag{4.12}$$

By similar arguments, one can show that

$$\bar{c} \le \alpha_t(D, \omega), \quad \omega \in P_i^t. \tag{4.13}$$

Since $\bar{\omega}$ was arbitrarily chosen, by (4.12) and (4.13), we finally conclude that,

$$\alpha_t(D, \omega) \ge m(\omega) \ge \alpha_t(D', \omega), \quad \text{for all} \quad \omega \in \Omega\,.$$

Thus (D7) is checked.

Let us show that α is right-continuous. Given $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$, we have

$$\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le 0\} \subseteq \{x \in (0, +\infty) : \rho_t^x(D, \omega) \le c\},$$

for any constant $c > 0$. Taking the supremum of both sides, and then the limit of the right-hand side as $c \to 0+$, we get

$$\sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le 0\} \le \lim_{c \to 0^+} \sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le c\}. \tag{4.14}$$

If the above inequality holds strictly, then there exists $x^0 \in (0, +\infty)$ such that

$$\sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le 0\} < x^0 < \lim_{c \to 0^+} \sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le c\}. \tag{4.15}$$

The second inequality implies that

$$x^0 < \sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le c\}, \quad \text{for all } c > 0.$$

By monotonicity of ρ^x, we deduce that $\rho_t^{x^0}(D, \omega) \le c$. Since the last inequality holds true for all $c > 0$, we have that $\rho_t^{x^0}(D, \omega) \le \lim_{c \to 0^+} c = 0$, that contradicts with first strict inequality in (4.15). Therefore, we have equality in (4.14). Using this equality, and (A6), translation invariance of ρ^x, we write

$$\begin{aligned}
\alpha_t(D, \omega) &= \sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le 0\} \\
&= \lim_{c \to 0^+} \sup\{x \in (0, +\infty) : \rho_t^x(D, \omega) \le c\} \\
&= \lim_{c \to 0^+} \sup\{x \in (0, +\infty) : \rho_t^x(D + c1_{\{t\}}, \omega) \le 0\} \\
&= \lim_{c \to 0^+} \alpha_t(D + c1_{\{t\}}, \omega),
\end{aligned}$$

and right continuity of α is established.

Finally, we will prove that α is normalized. Given a fixed $t \in \mathcal{T}$, consider the following cash positions

$$D_{pos} := 1_{\{t\}}, \quad D_{neg} := -1_{\{t\}}\,.$$

Recall that $\rho_t(0) = 0$. By (4.8) and (A6), we have

$$\begin{aligned}
\alpha_t(D_{pos}, \omega) &= \sup\{x \in (0, +\infty) : \rho_t^x(1_{\{t\}}, \omega) \le 0\} \\
&= \sup\{x \in (0, +\infty) : \rho_t^x(0, \omega) - 1 \le 0\} \\
&= \sup\{x \in (0, +\infty) : -1 \le 0\} = +\infty.
\end{aligned}$$

Similarly, one can show that $\alpha_t(D_{neg}, \omega) = 0$.

The proof is complete. $\square$

We conclude this section with the main result that provides a representation of a DCAI in terms of a family of DCRMs, and vise versa, a representation of DCRM in terms of a DCAI.

THEOREM 4.8.

(i) If α is a normalized, right-continuous, DCAI, then there exists a left-continuous and increasing family of DCRMs $(\rho^x)_{x\in(0,+\infty)}$, such that

$$\alpha_t(D,\omega) = \sup\left\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\right\}. \tag{4.16}$$

(ii) If $(\rho^x)_{x\in(0,+\infty)}$ is a left-continuous and increasing family of DCRMs, then there exists a right-continuous and normalized DCAI α such that,

$$\rho_t^x(D,\omega) := \inf\left\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}},\omega) \ge x\right\},$$

Here we assume that $\inf\emptyset = \infty$ *and* $\sup\emptyset = 0$.

Proof. (i) For every $x \in (0,+\infty)$, define $\rho^x = (\rho_t^x)_{t=0}^T$ as follows,

$$\rho_t^x(D,\omega) := \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}},\omega) \ge x\}, \tag{4.17}$$

for all $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$. By Theorem 4.6, $(\rho^x)_{x\in(0,+\infty)}$ is an increasing, left-continuous, family of DCRMs. We will show that

$$\alpha_t(D,\omega) = \sup\left\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\right\},$$

for all $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$.

For any $t \in \mathcal{T}$, $D \in \mathcal{D}$, $\omega \in \Omega$, and $y^{t,D,\omega} > \sup\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\}$, we have

$$\rho_t^{y^{t,D,\omega}}(D,\omega) > 0\,.$$

By (4.17) $\inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}},\omega) \ge y^{t,D,\omega}\} > 0$, and hence,

$$\alpha_t(D,\omega) = \alpha_t(D + 01_{\{t\}},\omega) < y^{t,D,\omega}\,.$$

Since the above inequality holds true for all $y^{t,D,\omega} > \sup\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\}$, we conclude that

$$\alpha_t(D,\omega) \le \sup\left\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\right\}.$$

Similarly, one can show that $\alpha_t(D,\omega) \ge \sup\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\}$.

(ii) Define the function α as follows,

$$\alpha_t(D,\omega) := \sup\left\{x \in (0,+\infty) : \rho_t^x(D,\omega) \le 0\right\}, \tag{4.18}$$

for all $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$. By Theorem 4.7, α is a right-continuous and normalized DCAI. Finally, one can check that

$$\rho_t^x(D,\omega) := \inf\{c \in \mathbb{R} : \alpha_t(D + c1_{\{t\}},\omega) \ge x\},$$

for all $x \in (0,+\infty)$, $t \in \mathcal{T}$, $D \in \mathcal{D}$ and $\omega \in \Omega$. □

5. SPECIAL CONSTRUCTION OF DCAIs

It is known that a DCRM has a representation similar to (2.2). One of the important discoveries done in the process of robust representation of dynamic risk measures, similar to (2.2), was that due to dynamic consistency property (A7), the set of probability measures $\mathcal{D}$ has to possess some additional features, which depend on how the dynamic consistency property (A7) is formulated. A set of probability measures having such additional features is referred to as a dynamically consistent set of probability measures (or, for short, a consistent set of probability measures).

In Section 5.1, we present our version of the concept of dynamically consistent set of probability measures, as well as some nontrivial examples of such sets. It is seen that our concept is different from the ones previously studied in the literature. Its form and properties have been dictated by the goal of using it in the context of robust representation of our DCAI.

In Section 5.2, we prove the representation theorem for DCRM in terms of consistent sets of probability measures. We conclude this section with representation theorem for DCAIs in terms of families of sequences of dynamically consistent sets of probability measures.

5.1. Dynamically Consistent Sequence of Sets of Probability Measures

In this section we discuss the concept of a dynamically consistent sequence of sets of probability measures.

In what follows we denote by $\mathcal{P}$ the set of all absolutely continuous probability measures with respect to the underlying probability $\mathbb{P}$, and $\mathcal{P}^e$ will stand for the set of all equivalent probability measures with respect to $\mathbb{P}$. Recall that our standing assumption is that $\mathbb{P}$ has full support. Note that in this case, due to the finiteness of Ω, the set $\mathcal{P}$ consists of all probability measures on Ω, and also $\mathcal{P}^e$ coincides with the set of all probability measures on Ω of full support.

DEFINITION 5.1. A sequence of sets of probability measures $\{\mathcal{Q}_t\}_{t=0}^T$, with $\mathcal{Q}_t \subseteq \mathcal{P}$, is called dynamically consistent with respect to the filtration $\mathbb{F}$, if the following inequalities hold true

(5.1)
$$1_A \min_{\omega\in A}\left\{\inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}[X\,|\,\mathcal{F}_{t+1}](\omega)\right\} \leq 1_A \inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}[X\,|\,\mathcal{F}_t] \leq 1_A \max_{\omega\in A}\left\{\inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}[X\,|\,\mathcal{F}_{t+1}](\omega)\right\},$$

for every $t \in \{0,\ldots,T-1\}$, $A \in \mathcal{F}_t$, and every random variable X.

DEFINITION 5.2. A set of probability measures $\mathcal{Q} \subseteq \mathcal{P}$ is called consistent with respect to filtration $\mathbb{F}$, if the following equality holds true

$$\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X\,|\,\mathcal{F}_t] = \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\Big[\inf_{M\in\mathcal{Q}} \mathbb{E}_M[X\,|\,\mathcal{F}_{t+1}]\,|\,\mathcal{F}_t\Big], \tag{5.2}$$

for every $t \in \{0,\ldots,T-1\}$, and every random variable X.

PROPOSITION 5.3. *If a set of probability measures $\mathcal{Q} \subseteq \mathcal{P}$ is consistent, then $\{\mathcal{Q}_t\}_{t=0}^T$ with $\mathcal{Q}_t = \mathcal{Q}$, $t \in \mathcal{T}$, is dynamically consistent.*

Proof. If $\mathcal{Q} \subseteq \mathcal{P}$ is strongly consistent, then, for every $A \in \mathcal{F}_t$, $\mathcal{F}_t$-measurable random variable X, and $t \in \{0, \dots, T-1\}$, we have

$$
\begin{aligned}
(5.3)\qquad 1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X|\mathcal{F}_t] &= 1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\Big[\inf_{M\in\mathcal{Q}} \mathbb{E}_M\big[X \mid \mathcal{F}_{t+1}\big]\big| \mathcal{F}_t\Big] \\
&= 1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\Big[1_A \inf_{M\in\mathcal{Q}} \mathbb{E}_M[X \mid \mathcal{F}_{t+1}]\big| \mathcal{F}_t\Big] \\
&\le 1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\Big[1_A \max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X \mid \mathcal{F}_{t+1}](\omega)\Big\}\big| \mathcal{F}_t\Big] \\
&\le 1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\Big[\max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X \mid \mathcal{F}_{t+1}](\omega)\Big\}\big| \mathcal{F}_t\Big].
\end{aligned}
$$

Since $\max_{\omega\in A}\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X \mid \mathcal{F}_{t+1}](\omega)\}$ is a constant, then, for each $Q \in \mathcal{Q}$, we have,

$$
\mathbb{E}_Q\Big[\max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\big[X \mid \mathcal{F}_{t+1}\big](\omega)\Big\} \mid \mathcal{F}_t\Big] = \max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\big[X \mid \mathcal{F}_{t+1}\big](\omega)\Big\}.
$$

Therefore,

$$
\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q\Big[\max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X \mid \mathcal{F}_{t+1}](\omega)\Big\} \mid \mathcal{F}_t\Big] = \max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X \mid \mathcal{F}_{t+1}](\omega)\Big\}.
$$

The last equality together with (5.3) imply

$$
1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X|\mathcal{F}_t] \le 1_A \max_{\omega\in A}\Big\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X \mid \mathcal{F}_{t+1}](\omega)\Big\}.
$$

Finally note that for any set of probability measures $\mathcal{Q} \subseteq \mathcal{P}$, we have

$$
(5.4)\qquad 1_A \min_{\omega\in A}\Big\{\inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{F}_{t+1}](\omega)\Big\} \le 1_A \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{F}_t],
$$

for every $t \in \{0, \dots, T-1\}$, $A \in \mathcal{F}_t$, and every random variable X. □

The rest of the subsection is dedicated to examples of dynamically consistent sequences of sets of probability measures.

EXAMPLE 5.4. Singleton set $\mathcal{Q} = \{Q\}$, with $\mathbb{Q} \in \mathcal{P}^e$, is clearly strongly consistent. By Proposition 5.3 the constant sequence$\{\mathcal{Q}, \mathcal{Q}, \dots, \mathcal{Q}\}$ is dynamically consistent. For simplicity of writing, we will denote this sequence by $\mathcal{Q}^s$.

EXAMPLE 5.5. It is not hard to show that

$$
\sum_{i=1}^{n_t} 1_{P_i^t} \inf_{\mathbb{Q}\in\mathcal{P}^e} \mathbb{E}_{\mathbb{Q}}[D \mid \mathcal{F}_t] = \sum_{i=1}^{n_t} 1_{P_i^t} \min_{\omega\in P_i^t} D(\omega), \qquad t \in \mathcal{T},\ D \in \mathcal{D}.
$$

This implies that the set $\mathcal{P}^e$ of all equivalent probability measures with respective to $\mathbb{P}$, is consistent. Hence, the constant sequence $\{\mathcal{P}^e, \mathcal{P}^e, \dots, \mathcal{P}^e\}$ is dynamically consistent.

EXAMPLE 5.6. Let $a \ge 1$ be a real number. The following set of probability measures

$$
\mathcal{Q}^{a,u} := \{\mathbb{Q} \in \mathcal{P}^e \mid \mathbb{E}_{\mathbb{P}}[d\mathbb{Q}/d\mathbb{P}|\mathcal{F}_t] \le a\mathbb{E}_{\mathbb{P}}[d\mathbb{Q}/d\mathbb{P}|\mathcal{F}_{t-1}] \text{ for all } t \in \{1, \dots, T\}\}
$$

is consistent.

First note that

$$\inf_{\widetilde{\mathbb{Q}}\in\mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[X\,|\,\mathcal{F}_t] \geq \inf_{\mathbb{Q}\in\mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{Q}}\big[\inf_{\mathbb{M}\in\mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{M}}[X\,|\,\mathcal{F}_{t+1}]|\mathcal{F}_t\big],$$

for every $t \in \{0,\dots,T-1\}$ and $\mathcal{F}_t$-measurable bounded random variable X. Next we will show that the converse inequality also holds true and hence, by definition, $\mathcal{Q}^{a,u}$ is consistent. Towards this end, assume that $t \in \mathcal{T}$, X is an $\mathcal{F}_t$-measurable random variable, and $a \geq 1$; all arbitrary but fixed in what follows. For convenience, we denote by $P_{i,j}^{t+1}$ the set of partition $(P_1^{t+1},\dots,P_{n_{t+1}}^{t+1})$ such that $P_i^t = \cup_{j=1}^{k_i} P_{i,j}^{t+1},\ i=1,\dots,n_t$. Note that $k_1+k_2+\ldots+k_{n_t} = n_{t+1}$.

Pick up arbitrarily $n_t + n_{t+1}$ probability measures from $\mathcal{Q}^{a,u}$, and denote them by $(\mathbb{Q}_1,\mathbb{Q}_2,\dots,\mathbb{Q}_{n_t},\mathbb{M}_{1,1},\ \mathbb{M}_{1,2},\dots,\mathbb{M}_{1,k_1},\mathbb{M}_{2,1},\mathbb{M}_{2,2},\dots,\mathbb{M}_{2,k_2},\dots\dots,\mathbb{M}_{n_t,1},\mathbb{M}_{n_t,2},\dots,\mathbb{M}_{n_t,k_{n_t}})$. Some of them are allowed to be the same. We will construct a new probability measure based on the above set of probabilities. For any $i\in\{1,2,\dots,n_t\}$, $j\in\{1,2,\dots,k_i\}$, and $\omega\in P_{i,j}^{t+1}$ we put

$$\mathbb{H}(\omega) := \frac{\mathbb{M}_{i,j}(\omega)}{\mathbb{M}_{i,j}\big(P_{i,j}^{t+1}\big)}\frac{\mathbb{Q}_i\big(P_{i,j}^{t+1}\big)}{\mathbb{Q}_i\big(P_i^t\big)}\mathbb{P}\big(P_i^t\big).$$

Note that $P_{i,j}^{t+1}$, $i\in\{1,2,\dots,n_t\}, j\in\{1,2,\dots,k_i\}$, is a partition of Ω, and hence $\mathbb{H}$ is well defined, and since all probability measures in $\mathcal{Q}$ are of full support, $\mathbb{H}(\omega)$ is finite for all $\omega\in\Omega$. It is also easy to show that $\mathbb{H}(\Omega)=1$, and thus $\mathbb{H}$ is a probability measure.

Next we will prove that $\mathbb{H}\in\mathcal{Q}^{a,u}$. On any set P_i^t, we have

$$\begin{aligned}
1_{P_i^t}\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_t\right] &= 1_{P_i^t}\sum_{j=1}^{k_i}\sum_{\omega\in P_{i,j}^{t+1}}\frac{\mathbb{H}(\omega)}{\mathbb{P}\big(P_i^t\big)}\\
&= 1_{P_i^t}\sum_{j=1}^{k_i}\sum_{\omega\in P_{i,j}^{t+1}}\frac{\mathbb{M}_{i,j}(\omega)}{\mathbb{M}_{i,j}\big(P_{i,j}^{t+1}\big)}\frac{\mathbb{Q}_i\big(P_{i,j}^{t+1}\big)}{\mathbb{Q}_i\big(P_i^t\big)}\frac{\mathbb{P}\big(P_i^t\big)}{\mathbb{P}\big(P_i^t\big)}\\
&= 1_{P_i^t}\sum_{j=1}^{k_i}\frac{\mathbb{M}_{i,j}\big(P_{i,j}^{t+1}\big)}{\mathbb{M}_{i,j}\big(P_{i,j}^{t+1}\big)}\frac{\mathbb{Q}_i\big(P_{i,j}^{t+1}\big)}{\mathbb{Q}_i\big(P_i^t\big)}\frac{\mathbb{P}\big(P_i^t\big)}{\mathbb{P}\big(P_i^t\big)} = 1_{P_i^t}\sum_{j=1}^{k_i}\frac{\mathbb{Q}_i\big(P_{i,j}^{t+1}\big)}{\mathbb{Q}_i\big(P_i^t\big)} = 1_{P_i^t}\,.
\end{aligned}$$

Thus, $\mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{H}}{d\mathbb{P}}|\mathcal{F}_t]=1$, and by tower property, for all $s\leq t$, we also have $\mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{H}}{d\mathbb{P}}|\mathcal{F}_s]=1$. Consequently, we get

$$\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_s\right] \leq a\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{s-1}\right],\quad \text{for all } s\leq t\,. \tag{5.5}$$

On the other hand, for any $P_{i,j}^{t+1}$, we have,

$$1_{P_{i,j}^{t+1}}\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{t+1}\right] = 1_{P_{i,j}^{t+1}}\sum_{\omega\in P_{i,j}^{t+1}}\frac{\mathbb{H}(\omega)}{\mathbb{P}\big(P_{i,j}^{t+1}\big)} = 1_{P_{i,j}^{t+1}}\frac{\mathbb{Q}_i\big(P_{i,j}^{t+1}\big)}{\mathbb{Q}_i\big(P_i^t\big)}\frac{\mathbb{P}\big(P_i^t\big)}{\mathbb{P}\big(P_{i,j}^{t+1}\big)}\,.$$

Since $\mathbb{Q}_i\in\mathcal{Q}^{a,u}$, we have that $\mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{Q}_i}{d\mathbb{P}}|\mathcal{F}_{t+1}]\leq a\mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{Q}_i}{d\mathbb{P}}|\mathcal{F}_t]$, and thus $1_{P_{i,j}^{t+1}}\mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{Q}_i}{d\mathbb{P}}|\mathcal{F}_{t+1}]\leq a1_{P_{i,j}^{t+1}}\mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{Q}_i}{d\mathbb{P}}|\mathcal{F}_t]$. This implies that $1_{P_{i,j}^{t+1}}\frac{\mathbb{Q}_i(P_{i,j}^{t+1})}{\mathbb{P}(P_{i,j}^{t+1})}\leq a1_{P_{i,j}^{t+1}}\frac{\mathbb{Q}_i(P_i^t)}{\mathbb{P}(P_i^t)}$. Hence, $\frac{\mathbb{Q}_i(P_{i,j}^{t+1})}{\mathbb{P}(P_{i,j}^{t+1})}\frac{\mathbb{P}(P_i^t)}{\mathbb{Q}_i(P_i^t)}\leq a$, and therefore,

$$1_{P_{i,j}^{t+1}}\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{t+1}\right] \le 1_{P_{i,j}^{t+1}}a = 1_{P_{i,j}^{t+1}}a\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{t}\right].$$

Since the above holds true for any $P_{i,j}^{t+1}$, we have that

$$\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{t+1}\right] \le a\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{t}\right].$$

By similar arguments as above, inductively, one can show that

$$\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{s}\right] \le a\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{H}}{d\mathbb{P}}\middle|\mathcal{F}_{t}\right],$$

for any $s > t$. Combining this with (5.5), we conclude that $\mathbb{H} \in \mathcal{Q}^{a,u}$.

Next let us evaluate $\mathbb{E}_{\mathbb{H}}[D \mid \mathcal{F}_t]$. Consider a new random variable Y, defined as follows:

$$Y := \sum_{i=1}^{n_t}\sum_{j=1}^{k_i} 1_{P_{i,j}^{t+1}}\mathbb{E}_{\mathbb{M}_{i,j}}[D \mid \mathcal{F}_{t+1}].$$

Then, for any $m \in \{1, 2, \dots, n_t\}$, we can deduce that

$$1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y \mid \mathcal{F}_t] = 1_{P_m^t}\sum_{j=1}^{k_m}\mathbb{E}_{\mathbb{Q}_m}[1_{P_{m,j}^{t+1}}\mathbb{E}_{\mathbb{M}_{m,j}}[D \mid \mathcal{F}_{t+1}]|\mathcal{F}_t]. \tag{5.6}$$

For convenience, we put $C_{m,j}^{t+1} := 1_{P_{m,j}^{t+1}}\sum_{\omega\in P_{m,j}^{t+1}}\frac{\mathbb{M}_{m,j}(\omega)D(\omega)}{\mathbb{M}_{m,j}(P_{m,j}^{t+1})}$. Note that

$$1_{P_{m,j}^{t+1}}\mathbb{E}_{\mathbb{M}_{m,j}}[D \mid \mathcal{F}_{t+1}] = C_{m,j}^{t+1}.$$

Hence,

$$1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y| \mathcal{F}_t] = 1_{P_m^t}\sum_{j=1}^{k_m}\mathbb{E}_{\mathbb{Q}_m}\left[C_{m,j}^{t+1}\middle|\mathcal{F}_t\right] = \sum_{\bar\omega\in P_m^t}\frac{\mathbb{Q}_m(\bar\omega)}{\mathbb{Q}_m(P_m^t)}\left(\sum_{j=1}^{k_m}C_{m,j}^{t+1}(\bar\omega)\right).$$

By the definition of $C_{m,j}^{t+1}$, we note that

$$\sum_{j=1}^{k_m}C_{m,j}^{t+1}(\bar\omega) = \sum_{j=1}^{k_m}\left[1_{P_{m,j}^{t+1}}(\bar\omega)\sum_{\omega\in P_{m,j}^{t+1}}\frac{\mathbb{M}_{m,j}(\omega)}{\mathbb{M}_{m,j}\left(P_{m,j}^{t+1}\right)}D(\omega)\right].$$

Then,

$$\begin{aligned} 1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y| \mathcal{F}_t] &= \sum_{\bar\omega\in P_m^t}\left[\frac{\mathbb{Q}_m(\bar\omega)}{\mathbb{Q}_m\left(P_m^t\right)}\sum_{j=1}^{k_m}\left[1_{P_{m,j}^{t+1}}(\bar\omega)\sum_{\omega\in P_{m,j}^{t+1}}\frac{\mathbb{M}_{m,j}(\omega)}{\mathbb{M}_{m,j}\left(P_{m,j}^{t+1}\right)}D(\omega)\right]\right] \\ &= \sum_{u=1}^{k_m}\sum_{\omega\in P_{m,u}^{t+1}}\frac{\mathbb{Q}_m\left(P_{m,u}^{t+1}\right)}{\mathbb{Q}_m\left(P_m^t\right)}\frac{\mathbb{M}_{m,u}(\omega)}{\mathbb{M}_{m,u}\left(P_{m,u}^{t+1}\right)}D(\omega). \end{aligned}$$

From here, using the fact that $\mathbb{H}(P_i^t) = \mathbb{P}(P_i^t)$, we conclude that

$$1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y \mid \mathcal{F}_t] = 1_{P_m^t}\mathbb{E}_{\mathbb{H}}[D \mid \mathcal{F}_t].$$

Since $\mathbb{H} \in \mathcal{Q}^{a,u}$, we have that $\mathbb{E}_{\mathbb{H}}[D \mid \mathcal{F}_t] \geq \inf_{\widetilde{\mathbb{Q}} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[D \mid \mathcal{F}_t]$. Consequently, the following inequality holds true

$$1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y \mid \mathcal{F}_t] \geq 1_{P_m^t} \inf_{\widetilde{\mathbb{Q}} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[D \mid \mathcal{F}_t].$$

By (5.6), it follows that

$$1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y \mid \mathcal{F}_t] = 1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}\left[\sum_{i=1}^{n_t}\sum_{j=1}^{k_i} 1_{P_{i,j}^{t+1}}\mathbb{E}_{\mathbb{M}_{i,j}}[D \mid \mathcal{F}_{t+1}]|\mathcal{F}_t\right],$$

from which we continue

$$\begin{aligned} 1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}[Y \mid \mathcal{F}_t] &\geq 1_{P_m^t}\mathbb{E}_{\mathbb{Q}_m}\left[\sum_{i=1}^{n_t}\sum_{j=1}^{k_i} 1_{P_{i,j}^{t+1}} \inf_{\mathbb{M}_{i,j} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{M}_{i,j}}[D \mid \mathcal{F}_{t+1}]\middle|\mathcal{F}_t\right] \\ &\geq 1_{P_m^t} \inf_{\widetilde{\mathbb{Q}} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[D \mid \mathcal{F}_t], \end{aligned}$$

and since this is true for all $\mathbb{Q}_m \in \mathcal{Q}^{x,u}$, we have

$$1_{P_m^t} \inf_{\mathbb{Q}_m \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{Q}_m}\left[\sum_{i=1}^{n_t}\sum_{j=1}^{k_i} 1_{P_{i,j}^{t+1}} \inf_{\mathbb{M}_{i,j} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{M}_{i,j}}[D \mid \mathcal{F}_{t+1}]\middle|\mathcal{F}_t\right] \geq 1_{P_m^t} \inf_{\widetilde{\mathbb{Q}} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[D \mid \mathcal{F}_t].$$

Summing both sides of the last inequality over $m \in \{1, 2, \ldots, n_t\}$, we have

$$\sum_{m=1}^{n_t} 1_{P_m^t} \inf_{\mathbb{Q}_m \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{Q}_m}\left[\sum_{i=1}^{n_t}\sum_{j=1}^{k_i} 1_{P_{i,j}^{t+1}} \inf_{\mathbb{M}_{i,j} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{M}_{i,j}}[D \mid \mathcal{F}_{t+1}]\middle|\mathcal{F}_t\right] \geq \sum_{m=1}^{n_t} 1_{P_m^t} \inf_{\widetilde{\mathbb{Q}} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[D \mid \mathcal{F}_t],$$

or equivalently,

$$\inf_{\mathbb{Q} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{Q}}\left[\inf_{\mathbb{M} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\mathbb{M}}[D \mid \mathcal{F}_{t+1}]|\mathcal{F}_t\right] \geq \inf_{\widetilde{\mathbb{Q}} \in \mathcal{Q}^{a,u}} \mathbb{E}_{\widetilde{\mathbb{Q}}}[D \mid \mathcal{F}_t].$$

This concludes the proof that $\mathcal{Q}^{a,u}$ is consistent.

REMARK 5.7. It is easy to show that for any $\mathbb{Q} \in \mathcal{Q}^{a,u}$,

$$\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\mathcal{F}_t\right] \leq a^t, \quad t \in \mathcal{T}.$$

In particular, $\mathbb{Q}(A) \leq a^t\mathbb{P}(A)$, for any $\mathbb{Q} \in \mathcal{Q}^{a,u}$, $A \in \mathcal{F}_t$ and $t \in \mathcal{T}$. Different probabilities in $\mathcal{Q}^{a,u}$ can be regarded as different opinions about the distribution of cash flows; the above inequality provides an upper bound of these probabilities in terms of the underlying probability $\mathcal{P}$.

EXAMPLE 5.8. By similar arguments as in previous examples, one can show that the set of probability measures $\mathcal{Q}^{a,l}$ defined as follows

$$\mathcal{Q}^{a,l} := \{\mathbb{Q} \in \mathcal{P}^e \mid \mathbb{E}_{\mathbb{Q}}[d\mathbb{P}/d\mathbb{Q} \mid \mathcal{F}_j] \leq a\mathbb{E}_{\mathbb{Q}}[d\mathbb{P}/d\mathbb{Q} \mid \mathcal{F}_{j-1}] \text{ for all } j = 1, \dots, T, \}$$

is a consistent set of probability measures.

EXAMPLE 5.9. In this example we construct a dynamically consistent sequence of sets of probability measures that is not a constant sequence of consistent sets of probability measures.

Let $\mathbb{P}_0, \mathbb{P}_1, \dots, \mathbb{P}_T$, be a sequence of probability measures in $\mathcal{P}^e$ such that $\mathbb{P}_i \neq \mathbb{P}_j$, for $i \neq j$. Consider the following sequence of sets of probability measures $\mathcal{Q}_t = \mathcal{P} \setminus \mathbb{P}_t,\ t = 0, 1, \dots, T$. It is easy to show that

$$\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{F}_t] = \inf_{\mathbb{Q}\in\mathcal{P}^e} \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{F}_t], \quad t \in \mathcal{T}. \tag{5.7}$$

This implies that $\{\mathcal{Q}_t\}_{t=0}^T$ is a dynamically consistent sequence of sets of probabilities measures. Clearly it is not a constant sequence.

5.2. Representation Theorem of DCRM

In this section we will present a representation theorem for DCRMs in terms of a dynamically consistent set of probabilities. These results combined with the results from Section 4.2 about duality between DCAI and DCRM will lead to a representation theorem for DCAIs.

THEOREM 5.10 (Representation Theorem for DCRM). *A function $\rho : \{0, 1, \dots, T\} \times \mathcal{D} \times \Omega \to \mathbb{R}$ is a DCRM if and only if there exists a dynamically consistent family of sets of probabilities $\mathcal{U} := \{\mathcal{Q}_s\}_{s=0}^T$ such that,*

$$\rho_t(D) = -\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s|\mathcal{F}_t\right], \quad \textit{for all} \quad t \in \mathcal{T},\ D \in \mathcal{D}. \tag{5.8}$$

Proof. **Sufficiency**. It is not hard to show that ρ defined in (5.8) is a DCRM. (A.1)–(A.6) are checked similarly as in existing literature (see for instance Riedel 2004), and for interest of saving space we will not check them here. We will show only that (A.7), dynamic consistency, is satisfied.

Since $\mathcal{U} = \{\mathcal{Q}_t\}_{t=0}^T$ is dynamically consistent, we have

$$\begin{aligned}
1_A\rho_t(D) &= -1_A \inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s|\mathcal{F}_t\right] \\
&\geq 1_A \min_{\omega\in A}\left\{- \inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t+1}^{T} D_s|\mathcal{F}_{t+1}\right](\omega) - D_t\right\} \\
&= 1_A \min_{\omega\in A}\{\rho_{t+1}(D, \omega) - D_t\},
\end{aligned}$$

for any $A \in \mathcal{F}_t,\ t \in \mathcal{T},\ D \in \mathcal{D}$, and $\mathcal{Q}_t \in \mathcal{U}$.

Similarly, one can show that $1_A\rho_t(D) = 1_A\max_{\omega\in A}\{\rho_{t+1}(D, \omega) - D_t\}$, for every $t \in \mathcal{T},\ D \in \mathcal{D},\ \mathcal{Q}_t \in \mathcal{U}$. Thus (A.7) is satisfied.

Necessity. The set $\mathcal{U}$ will be constructed explicitly. Fix a time $t \in \mathcal{T}$. Recall that $\{P_1^t, \dots, P_{n_t}^t\}$ denotes the partition of Ω that corresponds to $\mathcal{F}_t$. Also, we will denote by $\{P_{i,1}^{t,s}, \dots, P_{i,m_s}^{t,s}\}$ the partition of P_i^t generated by $\mathcal{F}_s$, for some future time $s \geq t$. Thus $P_i^t = \cup_{j=1}^{m_s} P_{i,j}^{t,s}$. Assume that P_i^t is fixed for some $i \in \{1, \dots, n_t\}$, and define the following probability space $(\Omega_i^t, 2^{\Omega_i^t}, \mathbb{P}^{\text{uni}})$ with

$$\Omega_i^t := \left\{ \left(s, P_{i,j}^{t,s}\right) : s \in \{t, t+1, \dots, T\} \text{ and } j \in \{1, 2, \dots, m_s\} \right\},$$

and $\mathbb{P}^{\text{uni}}(\omega) = 1/\text{card}(\Omega_i^t)$ for each $\omega \in \Omega_i^t$.

Let us denote by $\mathcal{X}(\Omega_i^t)$ the set of all random variables on Ω_i^t. There exists a one-to-one correspondence between $\mathcal{X}(\Omega_i^t)$ and the set $\mathcal{D}_i^t := \{D 1_{\{t,t+1,\dots,T\}} 1_{P_i^t} : \text{ for all } D \in \mathcal{D}\}$. The map can be defined as follows: for any $X \in \mathcal{X}(\Omega_i^t)$, put

$$D_s^X(\omega) := \begin{cases} X\left(\left(s, P_{i,j}^{t,s}\right)\right), & \text{if } \ s \geq t \quad \text{and} \quad \omega \in P_{i,j}^{t,s} \\ 0, & \text{otherwise,} \end{cases} \tag{5.9}$$

and vice versa, for any $D \in \mathcal{D}_i^t$, define

$$X^D\left(\left(s, P_{i,j}^{t,s}\right)\right) := D_s(\omega), \tag{5.10}$$

for $s \geq t, j \in \{1, 2, \dots, m_s\}$, and $\omega \in P_{i,j}^{t,s}$.

Consider the following map $\phi : \mathcal{X}(\Omega_i^t) \to \mathbb{R}$ with,

$$\phi(X) := \tfrac{1}{T-t+1} \rho_t(D^X, \omega), \quad \omega \in P_i^t. \tag{5.11}$$

We claim that ϕ is a static CRM, i.e., satisfies the properties (R1)–(R4) of Definition 2.3. Indeed, for any $X, Y \in \mathcal{X}(\Omega_i^t)$, such that $X \leq Y$, we have $D_s^X(\omega) \leq D_s^Y(\omega)$, for all $s \geq t$ and $\omega \in \Omega$. Then, by (A3), the monotonicity of ρ, we get $\rho_t(D^X, \omega) \geq \rho_t(D^Y, \omega)$, for $\omega \in \Omega$. Therefore, by (5.11), $\phi(X) \geq \phi(Y)$, i.e., ϕ satisfies (R1).

Note that for all $X \in \mathcal{X}(\Omega_i^t)$ and $\lambda \geq 0$, by (5.9), we have,

$$D_s^{\lambda X}(\omega) = \lambda X\left(\left(s, P_{i,j}^{t,s}\right)\right) = \lambda D_s^X(\omega),$$

for all $s \geq t$ and $\omega \in P_{i,j}^{t,s}$. From here, by (5.11) and using homogeneity of ρ, the homogeneity (R2) of ϕ follows.

Next we will show that ϕ satisfies (R3). For all $X \in \mathcal{X}(\Omega_i^t)$ and $k \in \mathbb{R}$, by (5.9), we have

$$D_s^{X+k}(\omega) = X\left(\left(s, P_{i,j}^{t,s}\right)\right) + k = D_s^X(\omega) + k,$$

for all $s \geq t$ and $\omega \in P_{i,j}^{t,s}$. Therefore, by (5.11) and (A6), translation invariance of ρ, we deduce

$$\begin{aligned} \phi(X+k) &= \frac{1}{T-t+1} \rho_t(D^X + k 1_{\{t,\dots,T\}}, \omega) \\ &= \frac{1}{T-t+1} (\rho_t(D^X, \omega) - (T-t+1)k) \\ &= \phi(X^D) - k, \end{aligned}$$

for all $X \in \mathcal{X}(\Omega_i^t)$.

To show that ϕ satisfies (R4), consider an $X \in \mathcal{X}(\Omega_i^t)$. By (5.9) $D_s^{X+Y}(\omega) = D_s^X(\omega) + D_s^Y(\omega)$, for all $s \geq t$ and $\omega \in P_{i,j}^{t,s}$, and therefore, by (5.11) and (A.5), subadditivity of ρ, we obtain

$$\begin{aligned}\phi(X+Y) &= \frac{1}{T-t+1}\rho_t(D^X + D^Y, \omega) \le \frac{1}{T-t+1}(\rho_t(D^X, \omega) + \rho_t(D^Y, \omega))\\ &= \phi(X) + \phi(Y).\end{aligned}$$

From all the above, we conclude that ϕ is a static CRM. By Theorem 2.4, representation of static CRMs, there exists $\mathcal{M}_i^t$, a set of absolutely continuous probability measures with respect to $\mathbb{P}^{\text{uni}}$ on Ω_i^t, such that

$$\phi(X) = -\inf_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}}[X].$$

By (5.11), we have

$$\frac{1}{T-t+1}\rho_t(D^X, \omega) = -\inf_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}}[X], \quad \omega \in P_i^t. \tag{5.12}$$

Since there is one-to-one map between $\mathcal{X}(\Omega_i^t)$ and $\mathcal{D}_i^t$, for any $D \in \mathcal{D}_i^t$, we also can write

$$\frac{1}{T-t+1}\rho_t(D, \omega) = -\inf_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}}[X^D]. \tag{5.13}$$

Fix a time $t^0 \in \{t, t+1, \dots, T\}$, and denote by $\widetilde{D}$ the process $1_{\{t^0\}}$. By (A.6)-translation invariance and (A.2)-independence of the past of ρ, it follows that $\rho_t(\widetilde{D}, \omega) = -1, \omega \in P_i^t$. Hence, by (5.13),

$$\inf_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}}[X^{\widetilde{D}}] = \frac{1}{T-t+1}. \tag{5.14}$$

Note that $\mathbb{E}_{\mathbb{M}}[X^{\widetilde{D}}] = \mathbb{M}(\{t^0\} \times P_i^t)$. Thus, (5.14) implies

$$\inf_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{M}\big(\{t^0\} \times P_i^t\big) = \frac{1}{T-t+1}.$$

Similarly, one can show that $\mathbb{E}_{\mathbb{M}}[X^{-\widetilde{D}}] = -\mathbb{M}(\{t^0\} \times P_i^t)$. Thus we derive that

$$\inf_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}}[X^{-\widetilde{D}}] = \inf_{\mathbb{M}\in\mathcal{M}_i^t} \big(-\mathbb{M}\big(\{t^0\} \times P_i^t\big)\big) = -\sup_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{M}\big(\{t^0\} \times P_i^t\big),$$

and consequently

$$\sup_{\mathbb{M}\in\mathcal{M}_i^t} \mathbb{M}\big(\{t^0\} \times P_i^t\big) = \frac{1}{T-t+1}.$$

This yields that

$$\mathbb{M}\big(\{t^0\} \times P_i^t\big) = \frac{1}{T-t+1}, \quad t^0 \in \{t, t+1, \dots, T\}. \tag{5.15}$$

For any $s \in \{t, t+1, \dots, T\}$, define $\mathbb{M}^s : \Omega_i^t \to \mathbb{R}$ as follows

$$\mathbb{M}^s\big((r, P_{i,j}^{t,r})\big) := \begin{cases}(T-t+1)\mathbb{M}\big((r, P_{i,j}^{t,r})\big), & \text{when } \ r = s \ \text{ and } \ j \in \{1, 2, \dots, m_r\}\\ 0, & \text{otherwise.}\end{cases}$$

It is straightforward to show that $\mathbb{M}^s$ is a probability measure on Ω_i^t for every $s \in \{t, t+1, \ldots, T\}$.

For all $D \in \mathcal{D}$, we can derive

$$\begin{aligned}
\sum_{s=t}^{T} \mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_s}] &= \sum_{s=t}^{T} \left(\sum_{r=t}^{T} \sum_{j=1}^{m_r} \mathbb{M}^s\big((r, P_{i,j}^{t,r})\big)(D_s 1_s)_r(\omega) \right), \quad \text{for some } \omega \in P_{i,j}^{t,r} \\
&= \sum_{s=t}^{T} \left(\sum_{j=1}^{m_r} \mathbb{M}^s\big((s, P_{i,j}^{t,s})\big) D_s(\omega) \right), \quad \text{for some } \quad \omega \in P_{i,j}^{t,r} \\
&= \sum_{s=t}^{T} \left(\sum_{j=1}^{m_r} (T-t+1)\mathbb{M}\big((s, P_{i,j}^{t,s})\big) D_s(\omega) \right), \ \text{for some } \omega \in P_{i,j}^{t,r} \\
&= (T-t+1)\mathbb{E}_{\mathbb{M}}[X^D].
\end{aligned}$$

Hence, by (5.13), we have

$$\rho_t(D, \omega) = -(T-t+1) \inf_{\mathbb{M} \in \mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}}[X^D] = -\inf_{\mathbb{M} \in \mathcal{M}_i^t} \sum_{s=t}^{T} \mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_s}], \quad \omega \in P_i^t. \tag{5.16}$$

Since ρ satisfies (A.6) and (A.7), we deduce that

$$\rho_s(D_s 1_{\{s\}} - D_s 1_{\{T\}}, \omega) = 0, \quad s \geq t, \quad D \in \mathcal{D}, \quad \omega \in P_i^t.$$

Thus, (5.13) and (5.16) imply

$$\begin{aligned}
-\inf_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{D_s 1_{\{T\}}}]) &= -\inf_{\mathbb{M} \in \mathcal{M}_i^t} \left[\sum_{r=t}^{T} \mathbb{E}_{\mathbb{M}^r}[X^{(D_s 1_{\{s\}} - D_s 1_{\{T\}})_r 1_r}] \right] \\
&= \rho_t(D_s 1_{\{s\}} - D_s 1_{\{T\}}, \omega) = 0\,.
\end{aligned}$$

Since the above equality holds true for all $D \in \mathcal{D}$, it also holds true for $-D$. Hence, we have

$$\inf_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{-D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{-D_s 1_{\{T\}}}]) = 0\,. \tag{5.17}$$

On the other hand, by (5.10), one gets

$$\inf_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{-D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{-D_s 1_{\{T\}}}]) = -\sup_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{D_s 1_{\{T\}}}])\,.$$

Thus,

$$\sup_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{D_s 1_{\{T\}}}]) = 0. \tag{5.18}$$

By (5.17) and (5.18) we conclude that

$$\sup_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{D_s 1_{\{T\}}}]) = 0 = \inf_{\mathbb{M} \in \mathcal{M}_i^t} (\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}] - \mathbb{E}_{\mathbb{M}^T}[X^{D_s 1_{\{T\}}}]),$$

and hence

$$\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}] = \mathbb{E}_{\mathbb{M}^T}[X^{D_s 1_{\{T\}}}] \tag{5.19}$$

for all $s \geq t$, and $\mathbb{M} \in \mathcal{M}_i^t$. Therefore, we can rewrite (5.16) as follows,

$$\begin{aligned}\rho_t(D,\omega) &= -\inf_{\mathbb{M}\in\mathcal{M}_i^t}\left[\sum_{s=t}^{T}\mathbb{E}_{\mathbb{M}^s}[X^{D_s 1_{\{s\}}}]\right]\\ &= -\inf_{\mathbb{M}\in\mathcal{M}_i^t}\left[\mathbb{E}_{\mathbb{M}^T}\left[\sum_{s=t}^{T}X^{D_s 1_{\{T\}}}\right]\right]\\ &= -\inf_{\mathbb{M}\in\mathcal{M}_i^t}\mathbb{E}_{\mathbb{M}^T}\big[X^{(\sum_{s=t}^{T}D_s)1_{\{T\}}}\big]\end{aligned} \tag{5.20}$$

for all $D \in \mathcal{D}$, and $\omega \in P_i^t$.

To summarize, for every P_i^t, $i = 1, \dots, n_t$, we constructed a set of probability measures $\mathcal{M}_i^t$ on Ω_i^t. Having these sets, we define $\mathcal{Q}_t$ as follows:

$$\begin{aligned}\mathcal{Q}_t := \Bigg\{\mathbb{Q}\in\mathcal{P}\ :\ &\text{there exists } \{\mathbb{M}_i\}_{i=1}^{n_t} \text{ such that, for all } i\in\{1,\dots,n_t\},\ j\in\big\{1,\dots,m_T^i\big\},\\ &\mathbb{M}_i\in\mathcal{M}_i^t \text{ and } \mathbb{Q}(\omega) = \frac{1}{n_t}\frac{1}{\mathcal{N}\big(P_{i,j}^{t,T}\big)}\mathbb{M}_i^T\big(\big(T,\,P_{i,j}^{t,T}\big)\big) \text{ for all } \omega\in P_{i,j}^{t,T}\Bigg\},\end{aligned}$$

where $\mathcal{N}(P)$ stands for cardinality of the set $P \subset \Omega$.

By direct evaluations, one can show that $\mathcal{Q}_t$, $t \in \mathcal{T}$, is a set of probability measures on Ω.

Next we will show that (5.8) is fulfilled. Note that, for all $\omega \in P_i^t$,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T}D_s\middle|\mathcal{F}_t\right](\omega) &= \sum_{\omega\in P_i^t}\left[\sum_{s=t}^{T}D_s(\omega)\frac{\mathbb{Q}(\omega)}{\mathbb{Q}\big(P_i^t\big)}\right]\\ &= \sum_{j=1}^{m_T^i}\sum_{\omega\in P_{i,j}^{t,T}}\left[\sum_{s=t}^{T}D_s(\omega)\frac{1}{\mathcal{N}\big(P_{i,j}^{t,T}\big)}\mathbb{M}_i^T\big(\big(T,\,P_{i,j}^{t,T}\big)\big)\right]\\ &= \sum_{j=1}^{m_T^i}\left[\sum_{s=t}^{T}D_s(\omega)\mathbb{M}_i^T\big(\big(T,\,P_{i,j}^{t,T}\big)\big)\right]\\ &= \mathbb{E}_{\mathbb{M}_i^T}\big[X^{\sum_{s=t}^{T}D_s 1_{\{T\}}}\big].\end{aligned}$$

If $\inf_{\mathbb{Q}\in\mathcal{Q}_t}\mathbb{E}_{\mathbb{Q}}[\sum_{s=t}^{T}D_s|\mathcal{F}_t](\omega) > \inf_{\mathbb{M}_i\in\mathcal{M}_i^t}\mathbb{E}_{\mathbb{M}_i^T}[X^{\sum_{s=t}^{T}D_s 1_{\{T\}}}]$, then there exists $\widetilde{\mathbb{M}}_i \in \mathcal{M}_i^t$ such that

$$\mathbb{E}_{\widetilde{\mathbb{M}}_i^T}\big[X^{\sum_{s=t}^{T}D_s 1_{\{T\}}}\big] < \inf_{\mathbb{Q}\in\mathcal{Q}_t}\mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T}D_s|\mathcal{F}_t\right](\omega). \tag{5.21}$$

However, for $\widetilde{\mathbb{Q}}$ constructed by $\widetilde{\mathbb{M}}_i$, as previously proved,

$$\mathbb{E}_{\widetilde{\mathbb{M}}_i^T}\big[X^{\sum_{s=t}^T D_s 1_{\{T\}}}\big] = \mathbb{E}_{\widetilde{\mathbb{Q}}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right](\omega) \geq \inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right](\omega),\ \ \omega \in P_i^t,$$

that contradicts (5.21). By the same arguments, one can show that the inequality

$$\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right](\omega) < \inf_{\mathbb{M}_i\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}_i^T}\big[X^{\sum_{s=t}^T D_s 1_{\{T\}}}\big],$$

can not hold true, and thus, we conclude that

$$\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right](\omega) = \inf_{\mathbb{M}_i\in\mathcal{M}_i^t} \mathbb{E}_{\mathbb{M}_i^T}\big[X^{\sum_{s=t}^T D_s 1_{\{T\}}}\big], \quad \omega \in P_i^t,$$

and by (5.20),

$$\rho_t(D) = -\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right].$$

To complete the proof we need to show that $\{\mathcal{Q}_s\}_{s=0}^T$ is a dynamically consistent sequence of sets of probability measures. Recall that by (A.7), dynamic consistency of ρ,

$$1_A\big(\min_{\omega\in A} \rho_{t+1}(D,\omega) - D_t\big) \leq 1_A\rho_t(D) \leq 1_A\big(\max_{\omega\in A} \rho_{t+1}(D,\omega) - D_t\big), \tag{5.22}$$

for all $D \in \mathcal{D}$ and $A \in \mathcal{F}_t$. Using this, we get

$$1_A\left(\min_{\omega\in A}\left\{-\inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t+1}^T D_s|\mathcal{F}_{t+1}\right](\omega)\right\} - D_t\right) \leq 1_A\left(-\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right]\right),$$

for any $D \in \mathcal{D}$ and $A \in \mathcal{F}_t$. Consequently, we obtain

$$1_A \max_{\omega\in A}\left\{\inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_{t+1}\right](\omega)\right\} \geq 1_A \inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right],\ \ D \in \mathcal{D},\ \ A \in \mathcal{F}_t. \tag{5.23}$$

Similarly, by (5.22)

$$1_A\left(\max_{\omega\in A}\left\{-\inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t+1}^T D_s|\mathcal{F}_{t+1}\right](\omega)\right\} - D_t\right) \geq 1_A\left(-\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right]\right)$$

and hence

$$1_A \min_{\omega\in A}\left\{\inf_{\mathbb{Q}\in\mathcal{Q}_{t+1}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_{t+1}\right](\omega)\right\} \leq 1_A \inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right],\ \ D \in \mathcal{D},\ \ A \in \mathcal{F}_t. \tag{5.24}$$

Combining (5.23) and (5.24) dynamic consistency of $\{\mathcal{Q}_t\}_{t=0}^T$ follows.

This completes the proof. □

REMARK 5.11. An interesting question is whether the sequence $\{\mathcal{Q}_s\}_{s=0}^T$ appearing in the representation (5.8) can be replaced with a constant sequence of sets of probability measures. The question is motivated by the following observation:

First note that for any set of probability measures $\mathcal{Q} \subseteq \mathcal{P}$, the following inequality holds true

$$1_A \min_{\omega\in A}\{\inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{F}_{t+1}](\omega)\} \leq 1_A \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{F}_t], \tag{5.25}$$

for every $t \in \{0, \dots, T-1\}$, $A \in \mathcal{F}_t$, and every random variable X. Thus, if the set $\mathcal{Q}$ additionally satisfies the following weak consistency condition

$$1_A \max_{\omega\in A}\{\inf_{Q\in\mathcal{Q}} \mathbb{E}_{Q}[X \mid \mathcal{F}_{t+1}](\omega)\} \geq 1_A \inf_{Q\in\mathcal{Q}} \mathbb{E}_{Q}[X|\mathcal{F}_t], \tag{5.26}$$

then the constant sequence $\mathcal{Q}_t = \mathcal{Q},\ t \in \mathcal{T}$, is dynamically consistent. Observe that in Example 6.2 we indeed have that

$$\rho_t(D) = -\inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right], \quad t \in \mathcal{T}, \quad D \in \mathcal{D},$$

where $\mathcal{Q} = \mathcal{P}^e$. Note however that $\mathcal{P}^e$ satisfies consistency condition (5.2) which is stronger than (5.26).

5.3. Representation of DCAIs

Having derived a representation theorem for DCRMs in terms of sets of probability measures, and having derived the duality between DCRM and DCAI we can present the final results of this paper: duality between DCAI and sets of probability measures.

DEFINITION 5.12. A family of sequences of sets of probability measures $(\mathcal{U}^x := (\mathcal{Q}_t^x)_{t=0}^T)_{x\in(0,+\infty)}$ is called increasing if $\mathcal{Q}_t^x \supseteq \mathcal{Q}_t^y$, for all $x \geq y > 0$ and $t \in \mathcal{T}$.

As a direct consequence of Theorems 5.10 and 4.7 we have the following results:

THEOREM 5.13. *Assume that* $(\mathcal{U}^x := (\mathcal{Q}_t^x)_{t=0}^T)_{x\in(0,+\infty)}$ *is an increasing family of dynamically consistent sequences of sets of probability measures. Then, the function* α : $\{0, 1, \dots, T\} \times \mathcal{D} \times \Omega \to [0, +\infty]$ *defined as follows,*

$$\alpha_t(D) = \sup\{x \in (0, +\infty) : \inf_{\mathbb{Q}\in\mathcal{Q}_t^x} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^T D_s|\mathcal{F}_t\right] \geq 0\}, \quad t \in \mathcal{T},\ D \in \mathcal{D}, \tag{5.27}$$

is a normalized and right-continuous DCAI.

THEOREM 5.14. *If* α *is a normalized and right-continuous DCAI, then there exists a family of dynamically consistent sequences of sets of probability measures* $(\mathcal{U}^x := (\mathcal{Q}_t^x)_{t=0}^T)_{x\in(0,+\infty)}$ *such that*

$$\alpha_t(D) = \sup\{x \in (0, +\infty) : \inf_{\mathbb{Q}\in\mathcal{Q}_t^x} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s|\mathcal{F}_t\right] \geq 0\}, \quad t \in \mathcal{T},\ D \in \mathcal{D}.$$

Here we adopt the usual convention that $\inf\emptyset = \infty$ *and* $\sup\emptyset = 0$.

REMARK 5.15. We want to mention that the static AI is a particular case of the DCAI developed in this paper and corresponds to $T = 1$. The same is true for the representation theorem for static AI in terms of family of sets of probability measures.

6. EXAMPLES

Theorem 5.13, besides being a fundamental theoretical result, can serve as basis for construction of DCAIs by means of constructing increasing sequences of dynamic sets of probability measures. Using this idea, we present here some abstract, nontrivial, examples of DCAIs.

EXAMPLE 6.1 (Dynamic Upper-Limit Ratio). Assume that h: $(0, +\infty) \to [0, +\infty)$ is an increasing function. Define $\acute{\mathcal{Q}}^x$ as follows,

$$\acute{\mathcal{Q}}^x := \left\{\mathbb{Q} \in \mathcal{P}|\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\mathcal{F}_j\right] \leq (1 + h(x))\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\mathcal{F}_{j-1}\right] \text{ for all } j = 1, \dots, T, \right\},$$

and let $\mathcal{U}^x := \{\acute{\mathcal{Q}}^x\}_{t=0}^T$. Note that $\acute{\mathcal{Q}}^x = \mathcal{Q}^{1+h(x),u},\ x \geq 0$, where $\mathcal{Q}^{a,u}$, $a \geq 1$, is defined in Example 5.6, and thus $\acute{\mathcal{Q}}^x$ is dynamically consistent for any $x > 0$. Also observe that monotonicity of h implies monotonicity of $\acute{\mathcal{Q}}^x$ with respect to x. Hence, by Theorem 5.13,

$$\alpha_t(D) = \sup\left\{x \in (0, +\infty) : \inf_{\mathbb{Q}\in\acute{\mathcal{Q}}^x} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s\middle|\mathcal{F}_t\right] \geq 0\right\}.$$

is a normalized and right-continuous DCAI. We call it dynamic upper-limit ratio.

EXAMPLE 6.2 (Dynamic Lower-Limit Ratio). Similarly, using Example 5.8, we consider $\dot{\mathcal{Q}}^x := \mathcal{Q}^{1+h(x),l}$, for some increasing, nonnegative function h. Then, $\mathcal{U}^x := \{\dot{\mathcal{Q}}^x\}_{t=0}^T$ is dynamically consistent, and by Theorem 5.13, the function α defined by (5.27) with $\mathcal{Q}_t^x = \dot{\mathcal{Q}}^x,\ x > 0$, is a normalized and right-continuous DCAI. We call it dynamic lower-limit ratio.

EXAMPLE 6.3 (Continuation of Example 5.9). In Example 5.9 we constructed a nonconstant dynamically consistent sequence of sets of probability measures. In view of (5.7) the corresponding family of risk measures satisfies

$$\rho_t(D) = -\inf_{\mathbb{Q}\in\mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s\middle|\mathcal{F}_t\right] = -\inf_{\mathbb{Q}\in\mathcal{P}^e} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s\middle|\mathcal{F}_t\right], \quad \text{for all } t \in \mathcal{T},\ D \in \mathcal{D}.$$

The point that we are making here is that the infimum over a time-dependent set $\mathcal{Q}_t$ can be replaced by the infimum over time-independent set $\mathcal{P}^e$ (see also Remark 5.11 in this regard).

EXAMPLE 6.4 (Dynamic GLR). GLR is a typical return-to-risk type of performance measure, very popular among practitioners. We recall that it is defined as the ratio of expectation of positive returns to expectation of negative returns: $\mathrm{GLR}(X) := \mathbb{E}(X)/\mathbb{E}(\max\{-X, 0\})$, if $\mathbb{E}[X] > 0$, and zero otherwise. As shown in Cherny and Madan (2009), GLR is a (static) coherent acceptability measure.

Here we present a dynamic version of GLR, denoted by dGLR, and defined as follows:

$$\mathrm{dGLR}_t(D) := \begin{cases} \dfrac{\mathbb{E}\left[\sum_{s=t}^{T} D_s | \mathcal{F}_t\right]}{\mathbb{E}[\left(\sum_{s=t}^{T} D_s\right)^{-} | \mathcal{F}_t]}, & \text{if } \mathbb{E}\left[\sum_{s=t}^{T} D_s \middle| \mathcal{F}_t\right] > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{6.1}$$

where $(\sum_{s=t}^{T} D_s)^{-} := \max\{-\sum_{s=t}^{T} D_s, 0\}$ and $t \in \mathcal{T},\ D \in \mathcal{D}$. Note that taking $T = 1$, dGLR becomes the static GLR.

We argue that *dGLR is a normalized and right-continuous DCAI.* Indeed, since $\mathrm{dGLR}(1_T) = +\infty$ and $\mathrm{dGLR}(-1_T) = 0$, we have that dGLR is normalized. Right-continuity follows from linearity of expectation and continuity of function $f(x) = x^{-}$. Adaptiveness (D1), and independence of the past (D2) of dGLR follow directly from the definition. Monotonicity (D3), scale invariance (D4), and quasi-concavity (D5) are verified as in static case with expectation replaced by conditional expectation (for details see Cherny and Madan 2009).

Since $\mathbb{E}(\sum_{l=t}^{T}(D + m1_{\{s\}})_l|\mathcal{F}_t) = \mathbb{E}(\sum_{l=t}^{T}(D + m1_{\{t\}})_l|\mathcal{F}_t)$, and $\mathbb{E}((\sum_{l=t}^{T}(D + m1_{\{s\}})_l)^{-} |\mathcal{F}_t) = \mathbb{E}((\sum_{l=t}^{T}(D + m1_{\{t\}})_l)^{-}|\mathcal{F}_t)$, for all $t \in \mathcal{T},\ D \in \mathcal{D}$, (D6), translation invariance, follows.

Finally we will prove that dGLR satisfies (D7), dynamic consistency property. Assume that m is an $\mathcal{F}_t$-measurable random variable, and $D \in \mathcal{D}$ such that $D_t \leq 0$ and $\mathrm{dGLR}_{t+1}(D) \leq m$. Assume that $m \neq +\infty$, and $\mathbb{E}[\sum_{s=t+1}^{T} D_s|\mathcal{F}_{t+1}] > 0$. By definition of dGLR, we have $\mathbb{E}(\sum_{s=t+1}^{T} D_s|\mathcal{F}_{t+1}) \leq m \cdot \mathbb{E}(\{\sum_{s=t+1}^{T} D_s\}^{-}|\mathcal{F}_{t+1})$, and since $D_t \leq 0$, we have

$$\begin{aligned} \mathbb{E}\left(\sum_{s=t}^{T} D_s|\mathcal{F}_t\right) &\leq \mathbb{E}\left(\mathbb{E}(\sum_{s=t+1}^{T} D_s|\mathcal{F}_{t+1})|\mathcal{F}_t\right) \\ &\leq m\mathbb{E}\left(\left\{\sum_{s=t+1}^{T} D_s\right\}^{-} |\mathcal{F}_t\right) \leq m\mathbb{E}\left(\left\{\sum_{s=t}^{T} D_s\right\}^{-} |\mathcal{F}_t\right), \end{aligned}$$

which implies that $\mathrm{dGLR}_t(D) \leq x$. If $m = +\infty$ or $\mathbb{E}[\sum_{s=t+1}^{T} D_s|\mathcal{F}_{t+1}] \leq 0$, then clearly $\mathrm{dGLR}_t(D) \leq m$.

Similarly, one can show that if $D_t \geq 0$, and $\mathrm{dGLR}_{t+1}(D) \geq m$, then $\mathrm{dGLR}_t(D) \geq m$.

Thus, we conclude that dGLR is a DCAI.

EXAMPLE 6.5 (Counterexample). Taking into account the general form of a dynamic AI (cf. (5.27)), and the general form of a static one (cf. (2.1)), the natural question arises: is it possible to dynamize a static CAI by taking the appropriate "conditional quantity" of the cumulative future cash flow? For example, to dynamize GLR, we consider the static GLR, and replaced in it the expectation with conditional expectation, and the terminal

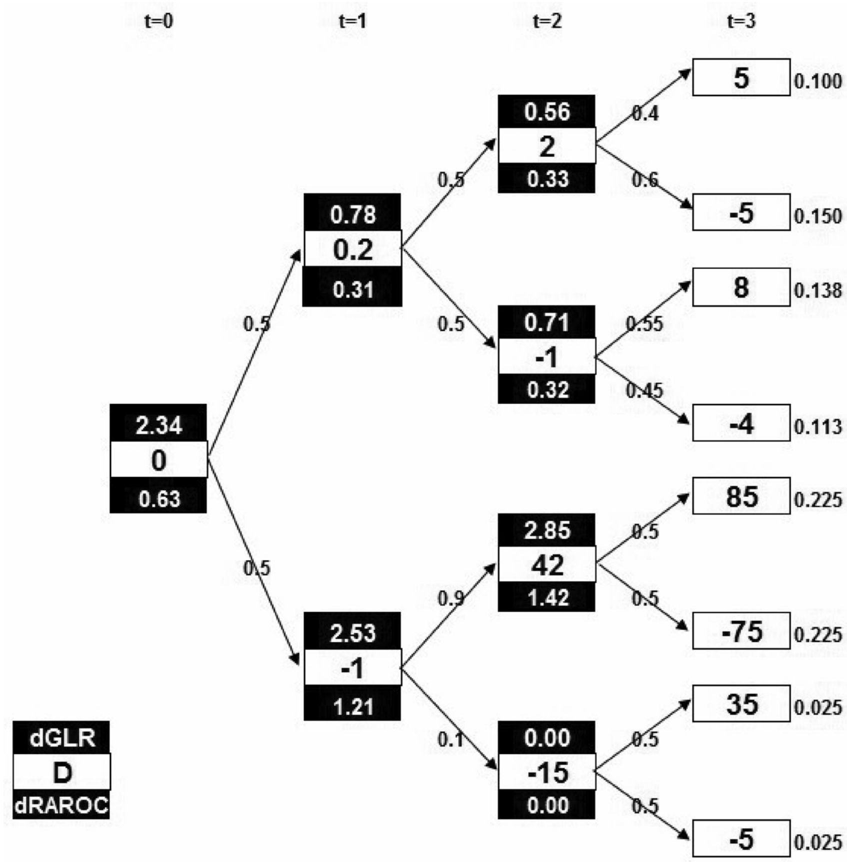


FIGURE 6.1. dRAROC vs. dGLR.

value with future cumulative cash flow. However, this procedure is not valid in general. The natural extension of static RAROC to a dynamic setup has the following form:

$$\mathrm{dRAROC}_t(D) = \begin{cases} \dfrac{\mathbb{E}\left(\sum_{s=t}^{T} D_s | \mathcal{F}_t\right)}{-\inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\left[\sum_{s=t}^{T} D_s | \mathcal{F}_t\right]}, & \text{when} \quad \mathbb{E}\left(\sum_{s=t}^{T} D_s | \mathcal{F}_t\right) > 0 \\ 0, & \text{otherwise} \end{cases}$$

with convention $\mathrm{dRAROC}_t(D) = +\infty$ if $\inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\sum_{s=t}^{T} D_s | \mathcal{F}_t] \geq 0$.

As is seen from Figure 6.1, which represents a numerical example, dRAROC does not satisfy property (D7), dynamic consistency. In this example, we consider $\mathcal{Q} = \mathcal{P}^e$. Assume that the states are labeled from top to bottom $\omega_1, \omega_2, \ldots, \omega_8$. Note that $D_1(\omega_1) = 0.2 > 0$, i.e., positive cash flow at time $t = 1$ and state ω_1, but $\mathrm{dRAROC}_1(\omega_1) = 0.31 < 0.33 = \mathrm{dRAROC}_2(\omega_1)$, as well as $\mathrm{dRAROC}_1(\omega_1) = 0.31 < 0.32 = \mathrm{dRAROC}_2(\omega_2)$. Thus dRAROC does not satisfy (D7) and hence it is not a DCAI.

For comparison reasons, we also present in Figure 6.1 the values of dGLR, which is a DCAI.

REFERENCES

ACCIAIO, B., and I. PENNER (2011): Dynamic Risk Measures, in *Advanced Mathematical Methods for Finance*, G. Di Nunno and B. Oksendal, eds., Heidelberg: Springer, pp 1–34.

ACCIAIO, B., H. FÖLLMER, and I. PENNER (2010): Risk Assessment for Uncertain Cash Flows: Model Ambiguity, Discounting Ambiguity, and the Role of Bubbles, Forthcoming in *Finance Stoch.*

ARTZNER, P., F. DELBAEN, J. M. EBER, and D. HEATH (1997): Thinking Coherently, *Risk* 10, 68–71.

ARTZNER, P., F. DELBAEN, J. M. EBER, and D. HEATH (1999): Coherent Measures of Risk, *Math. Finance* 9(3), 203–228.

ARTZNER, P., F. DELBAEN, J. M. EBER, D. HEATH, and H. KU (2007): Coherent Multiperiod Risk Adjusted Values and Bellman's Principle, *Ann. Oper. Res.* 152, 5–22.

BERNARDO, A., and O. LEDOIT (2000): Gain, Loss, and Asset Pricing, *J. Polit. Econ.* 108, 144–172.

BION-NADAL, J. (2009): Time Consistent Dynamic Risk Processes, *Stoch. Process. Appl.* 119(2), 633–654.

CARR, P., H. GEMAN, and D. B. MADAN (2001): Pricing and Hedging in Incomplete Markets, *J. Finan. Econ.* 62(1), 131–167.

CHERIDITO, P., and T. LI (2009): Risk Measures on Orlicz Hearts, *Math. Finance* 19(2), 189–214.

CHERIDITO, P., F. DELBAEN, and M. KUPPER (2004): Coherent and Convex Monetary Risk Measures for Bounded Càdlàg Processes, *Stoch. Process. Appl.* 112(1), 1–22.

CHERIDITO, P., F. DELBAEN, and M. KUPPER (2005): Coherent and Convex Monetary Risk Measures for Unbounded Càdlàg Processes, *Finance Stoch.* 9(3), 369–387.

CHERIDITO, P., F. DELBAEN, and M. KUPPER (2006): Dynamic Monetary Risk Measures for Bounded Discrete-time Processes, *Electron. J. Probab.* 11(3), 57–106.

CHERNY, A. S., and D. B. MADAN (2009): New Measures for Performance Evaluation, *Rev. Finan. Stud.* 22(7), 2571–2606.

CVITANIC, J., and I. KARATZAS (1999): On Dynamic Measures of Risk, *Finance Stoch.* 3(4), 451–482.

DELBAEN, F. (2000): *Coherent Risk Measures*, Pisa, Italy: Cattedra Galileiana, Scuola Normale Superiore.

DELBAEN, F. (2002): Coherent Risk Measures on General Probability Spaces, in *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, K. Sandmann and P. J. Schönbucher, eds. Berlin, Heidelberg, New York: Springer, pp. 1–37.

DELBAEN, F., S. PENG, and E. R. GIANIN (2010): Representation of the Penalty Term of Dynamic Concave Utilities, *Finance Stoch.* 14, 449–472.

DETLEFSEN, K., and G. SCANDOLO (2005): Conditional and Dynamic Convex Risk Measures, *Finance Stoch.* 9(4), 539–561.

FÖLLMER, H., and A. SCHIED (2002a): Convex Measures of Risk and Trading Constraints, *Finance Stoch.* 6(4), 429–447.

FÖLLMER, H., and A. SCHIED (2002b): Robust Preferences and Convex Measures of Risk, in *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, K. Sandmann and P. J. Schönbucher, eds. Berlin, Heidelberg, New York: Springer, pp. 39–56.

FÖLLMER, H., and A. SCHIED (2004): *Stochastic Finance, de Gruyter Studies in Mathematics*, extended edition: An Introduction in Discrete Time. Walter de Gruyter & Co., Berlin.

FRITTELLI, M., and E. R. GIANIN (2002): Putting Order in Risk Measures, *J. Bank. Finance* 26, 1473–1486.

FRITTELLI, M., and E. R. GIANIN (2004): Dynamic Convex Measures, in *New Risk Measures for the 21th Century*, G. Szegö, ed. Chichester: John Wiley & Sons, pp. 227–248.

FRITTELLI, M., and E. R. GIANIN (2005): Law Invariant Convex Risk Measures, *Adv. Math. Econ.* 7, 227–248.

FRITTELLI, M., and G. SCANDOLO (2006): Risk Measures and Capital Requirements for Processes, *Math. Finance* 16(4), 589–612.

JOBERT, A., and L. C. G. ROGERS (2008): Valuations and Dynamic Convex Risk Measures, *Math. Finance* 18(1), 1–22.

KUSUOKA, S. (2001): On Law Invariant Coherent Risk Measures, in *Advances in Mathematical Economics*, S. Kusuoka and T. Maruyama, eds. Tokyo: Springer, pp. 83–95.

RIEDEL, F. (2004): Dynamic Coherent Risk Measures, *Stoch. Process. Appl.* 112(2), 185–200.

ROORDA, B., J. M. SCHUMACHER, and J. ENGWERDA (2005): Coherent Acceptability Measures in Multiperiod Models, *Math. Finance* 15(4), 589–612.

SHARPE, W. F. (1966): Mutual Fund Performance, *J. Business* 39, 119–139.

SORTINO, F. A., and L. N. PRICE (1994): Performance Measurement in a Downside Risk Framework, *J. Invest.* 3(3), 59–64.

TUTSCH, S. (2008): Update Rules for Convex Risk Measures, *Quant. Finance* 8(8), 833–843.

WEBER, S. (2006): Distribution-invariant Risk Measures, Information, and Dynamic Consistency, *Math. Finance* 16(2), 419–441.

MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 2 (July 2014), 442–463

RISK MEASURES ON $\mathcal{P}(\mathbb{R})$ AND VALUE AT RISK WITH PROBABILITY/LOSS FUNCTION

MARCO FRITTELLI

Milano University

MARCO MAGGIS

Milano University

ILARIA PERI

ESC Rennes School of Business

We propose a generalization of the classical notion of the $V@R_\lambda$ that takes into account not only the probability of the losses, but the balance between such probability and the amount of the loss. This is obtained by defining a new class of law invariant risk measures based on an appropriate family of acceptance sets. The $V@R_\lambda$ and other known law invariant risk measures turn out to be special cases of our proposal. We further prove the dual representation of Risk Measures on $\mathcal{P}(\mathbb{R})$.

KEY WORDS: Value at Risk, distribution functions, quantiles, law invariant risk measures, quasi-convex functions, dual representation.

1. INTRODUCTION

We introduce a new class of law invariant risk measures $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ that are directly defined on the set $\mathcal{P}(\mathbb{R})$ of probability measures on $\mathbb{R}$ and are monotone and quasi-convex on $\mathcal{P}(\mathbb{R})$.

As Cherny and Madan (2009) pointed out, for a (*translation invariant*) coherent risk measure defined on random variables, all the positions can be spited in two classes: acceptable and not acceptable; in contrast, for an *acceptability index* there is a whole continuum of degrees of acceptability defined by a system $\{\mathcal{A}^m\}_{m\in\mathbb{R}}$ of sets. This formulation has been further investigated by Drapeau and Kupper (2010) for the quasi-convex case, with emphasis on the notion of an acceptability family and on the robust representation.

We adopt this approach and we build the maps Φ from a family $\{\mathcal{A}^m\}_{m\in\mathbb{R}}$ of acceptance sets of distribution functions by defining:

$$\Phi(P) := -\sup\{m \in \mathbb{R} \mid P \in \mathcal{A}^m\}.$$

The authors would like to thank Prof. Fabio Bellini, University Milano Bicocca, Samuel Drapeau, Humboldt University, as well as an anonymous referee for helpful discussion on this subject. MM acknowledges the financial support provided by the European Social Fund Grant.

Manuscript received February 2012; final revision received October 2012.

Address correspondence to Marco Frittelli, Department of Mathematics, Milano University, Via C. Saldini 50, Milano 20133, Italy; e-mail: Marco.Frittelli@unimi.it.

DOI: 10.1111/mafi.12028

In Section 3 we study the properties of such maps, we provide some specific examples, and in particular we propose an interesting generalization of the classical notion of $V@R_\lambda$.

The key idea of our proposal—the definition of the $\Lambda V@R$ in Section 4—arises from the consideration that to assess the risk of a financial position it is necessary to consider not only the probability λ of the loss, as in the case of the $V@R_\lambda$, but the dependence between such *probability* λ and the *amount* of the loss. In other terms, a risk prudent agent is willing to accept greater losses only with smaller probabilities. Hence, we replace the constant λ with a (increasing) function $\Lambda : \mathbb{R} \to [0, 1]$ defined on losses, which we call *Probability/Loss function*. The balance between the probability and the amount of the losses is incorporated in the definition of the family of acceptance sets

$$\mathcal{A}^m := \{Q \in \mathcal{P}(\mathbb{R}) \mid Q(-\infty, x] \leq \Lambda(x), \ \forall x \leq m\}, m \in \mathbb{R}.$$

If P_X is the distribution function of the random variable X, our new measure is defined by:

$$\Lambda V@R(P_X) := -\sup\{m \in \mathbb{R} \mid P(X \leq x) \leq \Lambda(x), \ \forall x \leq m\}.$$

As a consequence, the acceptance sets $\mathcal{A}^m$ are not obtained by the translation of $\mathcal{A}^0$ which implies that the map is not any more translation invariant. However, the similar property

$$\Lambda V@R(P_{X+\alpha}) = \Lambda^\alpha V@R(P_X) - \alpha,$$

where $\Lambda^\alpha(x) = \Lambda(x + \alpha)$, holds true and is discussed in Section 4.

The $V@R_\lambda$ and the worst case risk measure are special cases of the $\Lambda V@R$.

The approach of considering risk measures defined directly on the set of distribution functions is not new and it was already adopted by Weber (2006). However, in this paper we are interested in quasi-convex risk measures based—as the above mentioned map—on families of acceptance sets of distributions and in the analysis of their robust representation. We choose to define the risk measures on the entire set $\mathcal{P}(\mathbb{R})$ and not only on its subset of probabilities having compact support, as it was done by Drapeau and Kupper (2010). For this, we endow $\mathcal{P}(\mathbb{R})$ with the $\sigma(\mathcal{P}(\mathbb{R}), C_b(\mathbb{R}))$ topology. The selection of this topology is also justified by the fact (see Proposition 2.5) that for monotone maps $\sigma(\mathcal{P}(\mathbb{R}), C_b(\mathbb{R})) - lsc$ is equivalent to continuity from above. In Section 5 we briefly compare the robust representation obtained in this paper and those obtained by Cerreia-Vioglio (2009) and Drapeau and Kupper (2010).

Except for $\Phi = +\infty$, we show that there are no *convex*, $\sigma(\mathcal{P}(\mathbb{R}), C_b(\mathbb{R})) - lsc$ translation invariant maps $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$. But there are many quasi-convex and $\sigma(\mathcal{P}(\mathbb{R}), C_b(\mathbb{R})) - lsc$ maps $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ that in addition are monotone and translation invariant, as for example the $V@R_\lambda$, the entropic risk measure, and the worst case risk measure. This is another good motivation to adopt quasi-convexity versus convexity.

Finally, we provide the dual representation of quasi-convex, monotone, and $\sigma(\mathcal{P}(\mathbb{R}), C_b(\mathbb{R})) - lsc$ maps $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$—*defined on the entire set* $\mathcal{P}(\mathbb{R})$—and compute the dual representation of the risk measures associated to families of acceptance sets and consequently of the $\Lambda V@R$.

2. LAW INVARIANT RISK MEASURES

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $L^0 =: L^0(\Omega, \mathcal{F}, \mathbb{P})$ be the space of $\mathcal{F}$ measurable random variables that are $\mathbb{P}$ almost surely finite. Any random variable $X \in L^0$ induces a probability measure P_X on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ by $P_X(B) = \mathbb{P}(X^{-1}(B))$ for every Borel set $B \in \mathcal{B}_{\mathbb{R}}$. We refer to Aliprantis and Border (2005, chapter 15) for a detailed study of the convex set $\mathcal{P} =: \mathcal{P}(\mathbb{R})$ of probability measures on $\mathbb{R}$. Here we just recall some basic notions: for any $X \in L^0$ we have $P_X \in \mathcal{P}$ so that we will associate to any random variable a unique element in $\mathcal{P}$. If $\mathbb{P}(X = x) = 1$ for some $x \in \mathbb{R}$ then P_X is the Dirac distribution δ_x that concentrates the mass in the point x. A map $\rho : L \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$, defined on given subset $L \subseteq L^0$, is law invariant if $X, Y \in L$ and $P_X = P_Y$ implies $\rho(X) = \rho(Y)$.

Therefore, when considering law invariant risk measures $\rho : L^0 \to \overline{\mathbb{R}}$ it is natural to shift the problem to the set $\mathcal{P}$ by defining the new map $\Phi : \mathcal{P} \to \overline{\mathbb{R}}$ as $\Phi(P_X) = \rho(X)$. This map Φ is well defined on the entire $\mathcal{P}$, because there exists a bi-injective relation between $\mathcal{P}$ and the quotient space $\frac{L^0}{\sim}$ (provided that $(\Omega, \mathcal{F}, \mathbb{P})$ supports a random variable with uniform distribution), where the equivalence is given by $X \sim_{\mathcal{D}} Y \Leftrightarrow P_X = P_Y$. However, $\mathcal{P}$ is only a convex set and the usual operations on $\mathcal{P}$ are not induced by those on L^0, namely $(P_X + P_Y)(A) = P_X(A) + P_Y(A) \neq P_{X+Y}(A)$, $A \in \mathcal{B}_{\mathbb{R}}$.

Recall that the first-order stochastic dominance on $\mathcal{P}$ is given by: $Q \preccurlyeq P \Leftrightarrow F_P(x) \leq F_Q(x)$ for all $x \in \mathbb{R}$, where $F_P(x) = P(-\infty, x]$ and $F_Q(x) = Q(-\infty, x]$ are the distribution functions of $P, Q \in \mathcal{P}$. Note that $X \leq Y$ $\mathbb{P}$-a.s. implies $P_X \preccurlyeq P_Y$.

Definition 2.1. A Risk Measure on $\mathcal{P}(\mathbb{R})$ is a map $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ such that:

(Mon) Φ is monotone decreasing: $P \preccurlyeq Q$ implies $\Phi(P) \geq \Phi(Q)$;
(QCo) Φ is quasi-convex: $\Phi(\lambda P + (1-\lambda)Q) \leq \Phi(P) \vee \Phi(Q)$, $\lambda \in [0, 1]$.

Quasi-convexity can be equivalently reformulated in terms of sublevel sets: a map Φ is quasi-convex if for every $c \in \mathbb{R}$ the set $\mathcal{A}_c = \{P \in \mathcal{P} \mid \Phi(P) \leq c\}$ is convex. As recalled in Weber (2006), this notion of convexity is different from the one given for random variables (as in Föllmer and Schied 2004) because it does not concern diversification of financial positions. A natural interpretation in terms of compound lotteries is the following: whenever two probability measures P and Q are acceptable at some level c and $\lambda \in [0, 1]$ is a probability, then the compound lottery $\lambda P + (1 - \lambda)Q$, which randomizes over P and Q, is also acceptable at the same level. In terms of random variables (namely X, Y which induce P_X, P_Y), the randomized probability $\lambda P_X + (1-\lambda)P_Y$ will correspond to some random variable $Z \neq \lambda X + (1-\lambda)Y$ so that the diversification is realized at the level of distribution and not at the level of portfolio selection.

As suggested by Weber (2006), we define the translation operator T_m on the set $\mathcal{P}(\mathbb{R})$ by: $T_m P(-\infty, x] = P(-\infty, x - m]$, for every $m \in \mathbb{R}$. Equivalently, if P_X is the probability distribution of a random variable X we define the translation operator as $T_m P_X = P_{X+m}$, $m \in \mathbb{R}$. As a consequence we map the distribution $F_X(x)$ into $F_X(x - m)$. Note that $P \preccurlyeq T_m P$ for any $m > 0$.

Definition 2.2. If $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ is a risk measure on $\mathcal{P}$, we say that (TrI) Φ is translation invariant if $\Phi(T_m P) = \Phi(P) - m$ for any $m \in \mathbb{R}$.

Note that (TrI) corresponds exactly to the notion of cash additivity for risk measures defined on a space of random variables as introduced in Artzner et al. (1999). It is well known (see Cerreia-Vioglio et al. 2011b) that for maps defined on random variables, quasi-convexity and cash additivity imply convexity. However, in the context of distributions (QCo) and (TrI) do not imply convexity of the map Φ, as can be shown with the simple examples of the $V@R$ and the worst case risk measure ρ_w (see the examples in Section 3.1).

The set $\mathcal{P}(\mathbb{R})$ spans the space $ca(\mathbb{R}) := \{\mu \text{ signed measure} \mid V_\mu < +\infty\}$ of all signed measures of bounded variations on $\mathbb{R}$. $ca(\mathbb{R})$ (or simply ca) endowed with the norm $V_\mu = \sup\left\{\sum_{i=1}^{n} |\mu(A_i)| \text{ s.t. } \{A_1, \ldots, A_n\} \text{ partition of } \mathbb{R}\right\}$ is a norm complete and an Abstract Lebesgue space (see Aliprantis and Border 2005, paragraph 10.11).

Let $C_b(\mathbb{R})$ (or simply C_b) be the space of bounded continuous function $f : \mathbb{R} \to \mathbb{R}$. We endow $ca(\mathbb{R})$ with the weak* topology $\sigma(ca, C_b)$. The dual pairing $\langle\cdot,\cdot\rangle : C_b \times ca \to \mathbb{R}$ is given by $\langle f, \mu\rangle = \int f d\mu$ and the function $\mu \mapsto \int f d\mu$ ($\mu \in ca$) is $\sigma(ca, C_b)$ continuous. Note that $\mathcal{P}$ is a $\sigma(ca, C_b)$-closed convex subset of ca (p. 507 in Aliprantis and Border 2005) so that $\sigma(\mathcal{P}, C_b)$ is the relativization of $\sigma(ca, C_b)$ to $\mathcal{P}$ and any $\sigma(\mathcal{P}, C_b)$-closed subset of $\mathcal{P}$ is also $\sigma(ca, C_b)$-closed.

Even though $(ca, \sigma(ca, C_b))$ is not metrizable in general, its subset $\mathcal{P}$ is separable and metrizable (see Aliprantis and Border 2005, theorem 15.12) and therefore when dealing with convergence in $\mathcal{P}$ we may work with sequences instead of nets.

For every real function F we denote by $\mathcal{C}(F)$ the set of points in which the function F is continuous.

THEOREM 2.3 (Shiryaev 1995, theorem 2, p. 314). *Suppose that* $P_n, P \in \mathcal{P}$. *Then* $P_n \overset{\sigma(\mathcal{P},C_b)}{\longrightarrow} P$ *if and only if* $F_{P_n}(x) \to F_P(x)$ *for every* $x \in \mathcal{C}(F_P)$.

A sequence of probabilities $\{P_n\} \subset \mathcal{P}$ is decreasing, denoted with $P_n\downarrow$, if $F_{P_n}(x) \leq F_{P_{n+1}}(x)$ for all $x \in \mathbb{R}$ and all n.

DEFINITION 2.4. Suppose that $P_n, P \in \mathcal{P}$. We say that $P_n \downarrow P$ whenever $P_n \downarrow$ and $F_{P_n}(x) \uparrow F_P(x)$ for every $x \in \mathcal{C}(F_P)$. We say that (CfA) Φ *is continuous from above if* $P_n \downarrow P$ *implies* $\Phi(P_n) \uparrow \Phi(P)$.

PROPOSITION 2.5. *Let* $\Phi : \mathcal{P} \to \overline{\mathbb{R}}$ *be (Mon). Then the following are equivalent:*

Φ *is* $\sigma(\mathcal{P}, C_b)$*-lower semicontinuous*
Φ *is continuous from above.*

Proof. Let Φ be $\sigma(\mathcal{P}, C_b)$-lower semicontinuous and suppose that $P_n \downarrow P$. Then $F_{P_n}(x) \uparrow F_P(x)$ for every $x \in \mathcal{C}(F_P)$ and we deduce from Theorem 2.3 that $P_n \overset{\sigma(\mathcal{P},C_b)}{\longrightarrow} P$. (Mon) implies $\Phi(P_n) \uparrow$ and $k := \lim_n \Phi(P_n) \leq \Phi(P)$. The lower level set $A_k = \{Q \in \mathcal{P} \mid \Phi(Q) \leq k\}$ is $\sigma(\mathcal{P}, C_b)$ closed and, because $P_n \in A_k$, we also have $P \in A_k$, i.e., $\Phi(P) = k$, and Φ is continuous from above.

Conversely, suppose that Φ is continuous from above. As $\mathcal{P}$ is metrizable we may work with sequences instead of nets. For $k \in \mathbb{R}$ consider $A_k = \{P \in \mathcal{P} \mid \Phi(P) \leq k\}$ and a sequence $\{P_n\} \subseteq A_k$ such that $P_n \overset{\sigma(\mathcal{P},C_b)}{\longrightarrow} P \in \mathcal{P}$. We need to show that $P \in A_k$. Lemma 2.6 shows that each $F_{Q_n} := (\inf_{m\geq n} F_{P_m}) \wedge F_P$ is the distribution function of a probability measure and $Q_n \downarrow P$. From (Mon) and $P_n \preccurlyeq Q_n$, we get $\Phi(Q_n) \leq \Phi(P_n)$. From (CfA) then: $\Phi(P) = \lim_n \Phi(Q_n) \leq \liminf_n \Phi(P_n) \leq k$. Thus, $P \in A_k$. $\square$

LEMMA 2.6. *For every* $P_n \overset{\sigma(\mathcal{P},C_p)}{\longrightarrow} P$ *we have that*

$$F_{Q_n} := \inf_{m \geq n} F_{P_m} \wedge F_P, n \in \mathbb{N},$$

is a distribution function associated to a probability measure $Q_n \in \mathcal{P}$ *such that* $Q_n \downarrow P$.

Proof. For each n, F_{Q_n} is increasing and $\lim_{x\to-\infty} F_{Q_n}(x) = 0$. Moreover, for real valued maps right continuity and upper semicontinuity are equivalent. Because the inf-operator preserves upper semicontinuity we can conclude that F_{Q_n} is right continuous for every n. Now we have to show that for each n, $\lim_{x\to+\infty} F_{Q_n}(x) = 1$. By contradiction suppose that, for some n, $\lim_{x\to+\infty} F_{Q_n}(x) = \lambda < 1$. We can choose a sequence $\{x_k\}_k \subseteq \mathbb{R}$ with $x_k \in \mathcal{C}(F_P)$, $x_k \uparrow +\infty$. In particular, $F_{Q_n}(x_k) \leq \lambda$ for all k and $F_P(x_k) > \lambda$ definitively, say for all $k \geq k_0$. We can observe that because $x_k \in \mathcal{C}(F_P)$, we have, for all $k \geq k_0$, $\inf_{m\geq n} F_{P_m}(x_k) < \lim_{m\to+\infty} F_{P_m}(x_k) = F_P(x_k)$. This means that the infimum is attained for some index $m(k) \in \mathbb{N}$, i.e., $\inf_{m\geq n} F_{P_m}(x_k) = F_{P_{m(k)}}(x_k)$, for all $k \geq k_0$. Because $P_{m(k)}(-\infty, x_k] = F_{P_{m(k)}}(x_k) \leq \lambda$ then $P_{m(k)}(x_k, +\infty) \geq 1 - \lambda$ for $k \geq k_0$. We have two possibilities. Either the set $\{m(k)\}_k$ is bounded or $\overline{\lim}_k m(k) = +\infty$. In the first case, we know that the number of those $m(k)$ is finite. Among these $m(k)$ we can find at least one $\overline{m}$ and a subsequence $\{x_h\}_h$ of $\{x_k\}_k$ such that $x_h \uparrow +\infty$ and $P_{\overline{m}}(x_h, +\infty) \geq 1 - \lambda$ for every h. We then conclude that

$$\lim_{h\to+\infty} P_{\overline{m}}(x_h, +\infty) \geq 1 - \lambda$$

and this is a contradiction. If $\overline{\lim}_k m(k) = +\infty$, fix $\overline{k} \geq k_0$ such that $P(x_{\overline{k}}, +\infty) < 1 - \lambda$ and observe that for every $k > \overline{k}$

$$P_{m(k)}(x_{\overline{k}}, +\infty) \geq P_{m(k)}(x_k, +\infty) \geq 1 - \lambda.$$

Take a subsequence $\{m(h)\}_h$ of $\{m(k)\}_k$ such that $m(h) \uparrow +\infty$. Then:

$$\lim_{h\to\infty} \inf P_{m(h)}(x_{\overline{k}}, +\infty) \geq 1 - \lambda > P(x_{\overline{k}}, +\infty),$$

which contradicts the weak convergence $P_n \overset{\sigma(\mathcal{P},C_b)}{\longrightarrow} P$. Finally, note that $F_{Q_n} \leq F_{P_n}$ and $Q_n \downarrow$. From $P_n \overset{\sigma(\mathcal{P},C_b)}{\longrightarrow} P$ and the definition of Q_n, we deduce that $F_{Q_n}(x) \uparrow F_P(x)$ for every $x \in \mathcal{C}(F_P)$ so that $Q_n \downarrow P$. □

EXAMPLE 2.7 [The certainty equivalent]. It is very simple to build risk measures on $\mathcal{P}(\mathbb{R})$. Take any continuous, bounded from below and strictly decreasing function $f : \mathbb{R} \to \mathbb{R}$. Then the map $\Phi_f : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ defined by:

$$\Phi_f(P) := -f^{-1}\left(\int f dP\right) \tag{2.1}$$

is a Risk Measure on $\mathcal{P}(\mathbb{R})$. It is also easy to check that Φ_f is (CfA) and therefore $\sigma(\mathcal{P}, C_b)-lsc$. Note that Proposition 5.2 will then imply that Φ_f can not be convex. By selecting the function $f(x) = e^{-x}$ we obtain $\Phi_f(P) = \ln(\int \exp(-x) dF_P(x))$, which is in addition (TrI). Its associated risk measure $\rho : L^0 \to \mathbb{R} \cup \{+\infty\}$ defined on random variables, $\rho(X) = \Phi_f(P_X) = \ln(Ee^{-X})$, is the Entropic (convex) Risk Measure. In Section 5 we will see more examples based on this construction.

3. A REMARKABLE CLASS OF RISK MEASURES ON $\mathcal{P}(\mathbb{R})$

Given a family $\{F_m\}_{m\in\mathbb{R}}$ of functions $F_m : \mathbb{R} \to [0, 1]$, we consider the associated sets of probability measures

$$\mathcal{A}^m := \{Q \in \mathcal{P} \mid F_Q \leq F_m\} \tag{3.1}$$

and the associated map $\Phi : \mathcal{P} \to \overline{\mathbb{R}}$ defined by

$$\Phi(P) := -\sup\left\{m \in \mathbb{R} \mid P \in \mathcal{A}^m\right\}. \tag{3.2}$$

We assume hereafter that for each $P \in \mathcal{P}$ there exists m such that $P \notin \mathcal{A}^m$ so that $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$.

Note that $\Phi(P) := \inf\{m \in \mathbb{R} \mid P \in A_m\}$, where $A_m =: \mathcal{A}^{-m}$ and $\Phi(P)$ can be interpreted as the minimal risk acceptance level under which P is still acceptable. The following discussion will show that under suitable assumption on $\{F_m\}_{m\in\mathbb{R}}$ we have that $\{A_m\}_{m\in\mathbb{R}}$ is a risk acceptance family as defined in Drapeau and Kupper (2010).

We recall from Drapeau and Kupper (2010) the following definition

DEFINITION 3.1. A monotone decreasing family of sets $\{\mathcal{A}^m\}_{m\in\mathbb{R}}$ contained in $\mathcal{P}$ is *left continuous* in m if

$$\mathcal{A}^m =: \bigcap_{\varepsilon>0} \mathcal{A}^{m-\varepsilon}.$$

In particular it is *left continuous* if it is *left continuous* in m for every $m \in \mathbb{R}$.

LEMMA 3.2. *Let $\{F_m\}_{m\in\mathbb{R}}$ be a family of functions $F_m : \mathbb{R} \to [0, 1]$ and $\mathcal{A}^m$ be the set defined in (3.1). Then:*

1. *If, for every $x \in \mathbb{R}$, $F_\cdot(x)$ is decreasing (w.r.t. m) then the family $\{\mathcal{A}^m\}$ is monotone decreasing: $\mathcal{A}^m \subseteq \mathcal{A}^n$ for any level $m \geq n$.*
2. *For any m, $\mathcal{A}^m$ is convex and satisfies: $Q \preceq P \in \mathcal{A}^m \Rightarrow Q \in \mathcal{A}^m$.*
3. *If, for every $m \in \mathbb{R}$, $F_m(x)$ is right continuous w.r.t. x then $\mathcal{A}^m$ is $\sigma(\mathcal{P}, C_b)-$closed.*
4. *Suppose that, for every $x \in \mathbb{R}$, $F_m(x)$ is decreasing w.r.t. m. If $F_m(x)$ is left continuous w.r.t. m, then the family $\{\mathcal{A}^m\}$ is left continuous.*
5. *Suppose that, for every $x \in \mathbb{R}$, $F_m(x)$ is decreasing w.r.t. m and that, for every $m \in \mathbb{R}$, $F_m(x)$ is right continuous and increasing w.r.t. x and $\lim_{x\to+\infty} F_m(x) = 1$. If the family $\{\mathcal{A}^m\}$ is left continuous in m* then *$F_m(x)$ is left continuous in m.*

Proof.

1. If $Q \in \mathcal{A}^m$ and $m \geq n$ then $F_Q \leq F_m \leq F_n$, i.e., $Q \in \mathcal{A}^n$.
2. Let $Q, P \in \mathcal{A}^m$ and $\lambda \in [0, 1]$. Consider the convex combination $\lambda Q + (1-\lambda)P$ and note that

$$F_{\lambda Q+(1-\lambda)P} \leq F_Q \vee F_P \leq F_m,$$

as $F_P \leq F_m$ and $F_Q \leq F_m$. Then $\lambda Q + (1-\lambda)P \in \mathcal{A}^m$.

3. Let $Q_n \in A^m$ and $Q \in \mathcal{P}$ satisfy $Q_n \xrightarrow{\sigma(\mathcal{P},C_b)} Q$. By Theorem 2.3 we know that $F_{Q_n}(x) \to F_Q(x)$ for every $x \in \mathcal{C}(F_Q)$. For each n, $F_{Q_n} \leq F_m$ and therefore $F_Q(x) \leq F_m(x)$ for every $x \in \mathcal{C}(F_Q)$. By contradiction, suppose that $Q \notin \mathcal{A}^m$. Then

there exists $\bar{x} \notin \mathcal{C}(F_Q)$ such that $F_Q(\bar{x}) > F_m(\bar{x})$. By right continuity of F_Q for every $\varepsilon > 0$ we can find a right neighborhood $[\bar{x}, \bar{x} + \delta(\varepsilon))$ such that

$$|F_Q(x) - F_Q(\bar{x})| < \varepsilon \quad \forall\, x \in [\bar{x}, \bar{x} + \delta(\varepsilon))$$

and we may require that $\delta(\varepsilon) \downarrow 0$ if $\varepsilon \downarrow 0$. Note that for each $\varepsilon > 0$ we can always choose $x_\varepsilon \in (\bar{x}, \bar{x} + \delta(\varepsilon))$ such that $x_\varepsilon \in \mathcal{C}(F_Q)$. For such x_ε we deduce that

$$F_m(\bar{x}) < F_Q(\bar{x}) < F_Q(x_\varepsilon) + \varepsilon \leq F_m(x_\varepsilon) + \varepsilon.$$

This leads to a contradiction because if $\varepsilon \downarrow 0$ we have that $x_\varepsilon \downarrow \bar{x}$ and thus by right continuity of F_m:

$$F_m(\bar{x}) < F_Q(\bar{x}) \leq F_m(\bar{x}).$$

4. By assumption we know that $F_{m-\varepsilon}(x) \downarrow F_m(x)$ as $\varepsilon \downarrow 0$, for all $x \in \mathbb{R}$. By item 1, we know that $\mathcal{A}^m \subseteq \bigcap_{\varepsilon>0} \mathcal{A}^{m-\varepsilon}$. By contradiction, we suppose that the strict inclusion

$$\mathcal{A}^m \subset \bigcap_{\varepsilon>0} \mathcal{A}^{m-\varepsilon}$$

holds, so that there will exist $Q \in \mathcal{P}$ such that $F_Q \leq F_{m-\varepsilon}$ for every $\varepsilon > 0$ but $F_Q(\overline{x}) > F_m(\overline{x})$ for some $\overline{x} \in \mathbb{R}$. Set $\delta = F_Q(\overline{x}) - F_m(\overline{x})$ so that $F_Q(\overline{x}) > F_m(\overline{x}) + \frac{\delta}{2}$. Because $F_{m-\varepsilon} \downarrow F_m$ we may find $\overline{\varepsilon} > 0$ such that $F_{m-\overline{\varepsilon}}(\overline{x}) - F_m(\overline{x}) < \frac{\delta}{2}$. Thus, $F_Q(\overline{x}) \leq F_{m-\varepsilon}(\overline{x}) < F_m(\overline{x}) + \frac{\delta}{2}$ and this is a contradiction.
5. Assume that $\mathcal{A}^{m-\varepsilon} \downarrow \mathcal{A}^m$. Define $F(x) := \lim_{\varepsilon \downarrow 0} F_{m-\varepsilon}(x) = \inf_{\varepsilon>0} F_{m-\varepsilon}(x)$ for all $x \in \mathbb{R}$. Then $F : \mathbb{R} \to [0, 1]$ is increasing, right continuous (because the inf preserves this property). Note that for every $\varepsilon > 0$ we have $F_{m-\varepsilon} \geq F \geq F_m$ and then $\mathcal{A}^{m-\varepsilon} \supseteq \{Q \in \mathcal{P} \mid F_Q \leq F\} \supseteq \mathcal{A}^m$ and $\lim_{x \to +\infty} F(x) = 1$. Necessarily we conclude $\{Q \in \mathcal{P} \mid F_Q \leq F\} = \mathcal{A}^m$. By contradiction, we suppose that $F(\overline{x}) > F_m(\overline{x})$ for some $\overline{x} \in \mathbb{R}$. Define $F_{\overline{Q}} : \mathbb{R} \to [0, 1]$ by: $F_{\overline{Q}}(x) = F(x)\mathbf{1}_{[\overline{x},+\infty)}(x)$. The above properties of F guarantees that $F_{\overline{Q}}$ is a distribution function of a corresponding probability measure $\overline{Q} \in \mathcal{P}$, and because $F_{\overline{Q}} \leq F$, we deduce $\overline{Q} \in \mathcal{A}^m$, but $F_{\overline{Q}}(\overline{x}) > F_m(\overline{x})$ and this is a contradiction. □

The following Lemma can be deduced directly from the above Lemma 3.2 and from theorem 1.7 in Drapeau and Kupper (2010) (using the risk acceptance family $A_m =: \mathcal{A}^{-m}$, according to definition 1.6 in the aforementioned paper). We provide the proof for sake of completeness.

LEMMA 3.3. *Let $\{F_m\}_{m\in\mathbb{R}}$ be a family of functions $F_m : \mathbb{R} \to [0, 1]$ and Φ be the associated map defined in (3.2). Then:*

1. *The map Φ is (Mon) on $\mathcal{P}$.*
2. *If, for every $x \in \mathbb{R}$, $F_\cdot(x)$ is decreasing (w.r.t. m) then Φ is (QCo) on $\mathcal{P}$.*
3. *If, for every $x \in \mathbb{R}$, $F_\cdot(x)$ is left continuous and decreasing (w.r.t. m) and if, for every $m \in \mathbb{R}$, $F_m(\cdot)$ is right continuous (w.r.t. x) then*

$$A_m := \{Q \in \mathcal{P} \mid \Phi(Q) \leq m\} = \mathcal{A}^{-m}, \forall m, \tag{3.3}$$

and Φ is $\sigma(\mathcal{P}, C_b)$ lower semicontinuous.

Proof.

1. From $P \preccurlyeq Q$ we have $F_Q \leq F_P$ and

$$\{m \in \mathbb{R} \mid F_P \leq F_m\} \subseteq \left\{m \in \mathbb{R} \mid F_Q \leq F_m\right\},$$

which implies $\Phi(Q) \leq \Phi(P)$.

2. We show that $Q_1, Q_2 \in \mathcal{P}$, $\Phi(Q_1) \leq n$, and $\Phi(Q_2) \leq n$ imply that $\Phi(\lambda Q_1 + (1-\lambda)Q_2) \leq n$, that is

$$\sup\left\{m \in \mathbb{R} \mid F_{\lambda Q_1+(1-\lambda)Q_2} \leq F_m\right\} \geq -n.$$

By definition of the supremum, $\forall \varepsilon > 0$ $\exists m_i$ s.t. $F_{Q_i} \leq F_{m_i}$ and $m_i > -\Phi(Q_i) - \varepsilon \geq -n-\varepsilon$. Then $F_{Q_i} \leq F_{m_i} \leq F_{-n-\varepsilon}$, as $\{F_m\}$ is a decreasing family. Therefore, $\lambda F_{Q_1} + (1-\lambda)F_{Q_2} \leq F_{-n-\varepsilon}$ and $-\Phi(\lambda Q_1 + (1-\lambda)Q_2\lambda) \geq -n-\varepsilon$. As this holds for any $\varepsilon > 0$, we conclude that Φ is quasi-convex.

3. The fact that $\mathcal{A}^{-m} \subseteq A_m$ follows directly from the definition of Φ, as if $Q \in \mathcal{A}^{-m}$

$$\Phi(Q) := -\sup\left\{n \in \mathbb{R} \mid Q \in \mathcal{A}^n\right\} = \inf\left\{n \in \mathbb{R} \mid Q \in \mathcal{A}^{-n}\right\} \leq m.$$

We have to show that $A_m \subseteq \mathcal{A}^{-m}$. Let $Q \in A_m$. Because $\Phi(Q) \leq m$, for all $\varepsilon > 0$ there exists m_0 such that $m + \varepsilon > -m_0$ and $F_Q \leq F_{m_0}$. Because $F_{\cdot}(x)$ is decreasing (w.r.t. m) we have that $F_Q \leq F_{-m-\varepsilon}$, therefore, $Q \in \mathcal{A}^{-m-\varepsilon}$ for any $\varepsilon > 0$. By the left continuity in m of $F_{\cdot}(x)$, we know that$\{\mathcal{A}^m\}$ is left continuous (Lemma 3.2, item 4) and so: $Q \in \bigcap_{\epsilon>0} \mathcal{A}^{-m-\varepsilon} = \mathcal{A}^{-m}$.

From the assumption that $F_m(\,\cdot\,)$ is right continuous (w.r.t. x) and Lemma 3.2 (item 3), we already know that $\mathcal{A}^m$ is $\sigma(\mathcal{P}, C_b)$-closed, for any $m \in \mathbb{R}$, and therefore the lower level sets $A_m = \mathcal{A}^{-m}$ are $\sigma(\mathcal{P}, C_b)$-closed and Φ is $\sigma(\mathcal{P}, C_b)$ -lower-semicontinuous. □

DEFINITION 3.4. A family $\{F_m\}_{m\in\mathbb{R}}$ of functions $F_m : \mathbb{R} \to [0, 1]$ is *feasible* if

- For any $P \in \mathcal{P}$ there exists m such that $P \notin \mathcal{A}^m$.
- For every $m \in \mathbb{R}$, $F_m(\,\cdot\,)$ is right continuous (w.r.t. x).
- For every $x \in \mathbb{R}$, $F_{\cdot}(x)$ is decreasing and left continuous (w.r.t. m).

From Lemmas 3.2 and 3.3 we immediately deduce:

PROPOSITION 3.5. *Let $\{F_m\}_{m\in\mathbb{R}}$ be a feasible family. Then the associated family $\{\mathcal{A}^m\}_{m\in\mathbb{R}}$ is monotone decreasing and left continuous and each set $\mathcal{A}^m$ is convex and $\sigma(\mathcal{P}, C_b)$-closed. The associated map $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ is well defined, (Mon), (Qco), and $\sigma(\mathcal{P}, C_b)$-lsc.*

REMARK 3.6. Let $\{F_m\}_{m\in\mathbb{R}}$ be a feasible family. If there exists an $\overline{m}$ such that $\lim_{x\to+\infty} F_{\overline{m}}(x) < 1$ then $\lim_{x\to+\infty} F_m(x) < 1$ for every $m \geq \overline{m}$ and then $\mathcal{A}^m = \varnothing$ for every $m \geq \overline{m}$. Obviously, if an acceptability set is empty then it does not contribute to the computation of the risk measure defined in (3.2). For this reason we will always consider without loss of generality (w.l.o.g.) a class $\{F_m\}_{m\in\mathbb{R}}$ such that $\lim_{x\to+\infty} F_m(x) = 1$ for every m.

3.1. Examples

As explained in the introduction, we define a family of risk measures employing a Probability/Loss function Λ. Fix the *right continuous* function $\Lambda : \mathbb{R} \to [0, 1]$ and define the family $\{F_m\}_{m\in\mathbb{R}}$ of functions $F_m : \mathbb{R} \to [0, 1]$ by

$$F_m(x) := \Lambda(x)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x). \tag{3.4}$$

It is easy to check that if $\sup_{x\in\mathbb{R}} \Lambda(x) < 1$ then the family $\{F_m\}_{m\in\mathbb{R}}$ is feasible and therefore, by Proposition 3.5, the associated map $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ is well defined, (Mon), (Qco), and $\sigma(\mathcal{P}, C_b)$-*lsc*.

EXAMPLE 3.7. When $\sup_{x\in R} \Lambda(x) = 1$, Φ may take the value $-\infty$. The extreme case is when, in the definition of the family (3.4), the function Λ is equal to the constant one, $\Lambda(x) = 1$, and so: $\mathcal{A}^m = \mathcal{P}$ for all m and $\Phi = -\infty$.

EXAMPLE 3.8. Worst case risk measure: $\Lambda(x) = 0$.
Take in the definition of the family (3.4) the function Λ to be equal to the constant zero: $\Lambda(x) = 0$. Then:

$$\begin{aligned} F_m(x) &:= \mathbf{1}_{[m,+\infty)}(x), \\ \mathcal{A}^m &:= \left\{Q \in \mathcal{P} \mid F_Q \le F_m\right\} = \{Q \in \mathcal{P} \mid \delta_m \preccurlyeq Q\}, \\ \Phi_w(P) &:= -\sup\{m \mid P \in \mathcal{A}^m\} = -\sup\{m \mid \delta_m \preccurlyeq P\} \\ &= -\sup\{x \in \mathbb{R} \mid F_P(x) = 0\}, \end{aligned}$$

so that, if $X \in L^0$ has distribution function P_X,

$$\Phi_w(P_X) = -\sup\{m \in \mathbb{R} \mid \delta_m \preccurlyeq P_X\} = -ess\inf(X) := \rho_w(X)$$

coincide with the worst case risk measure ρ_w. As the family $\{F_m\}$ is feasible, $\Phi_w : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ is (Mon), (Qco), and $\sigma(\mathcal{P}, C_b)$-*lsc*. In addition, it also satisfies (TrI).

Even though ρ_w: $L^0 \to \mathbb{R} \cup \{\infty\}$ is convex, as a map defined on random variables, the corresponding $\Phi_w : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$, as a map defined on distribution functions, is not convex, but it is quasi-convex and quasi-concave. Indeed, let $P \in \mathcal{P}$ and, because $F_P \ge 0$, we set:

$$-\Phi_w(P) = \inf(F_P) := \sup\{x \in \mathbb{R} : F_P(x) = 0\}.$$

If F_1, F_2 are two distribution functions corresponding to P_1, $P_2 \in \mathcal{P}$ then for all $\lambda \in (0, 1)$ we have:

$$\inf(\lambda F_1 + (1-\lambda)F_2) = \min(\inf(F_1), \inf(F_2)) \le \lambda \inf(F_1) + (1-\lambda)\inf(F_2)$$

and therefore, for all $\lambda \in [0, 1]$

$$\min(\inf(F_1), \inf(F_2)) \le \inf(\lambda F_1 + (1-\lambda)F_2) \le \lambda \inf(F_1) + (1-\lambda)\inf(F_2).$$

EXAMPLE 3.9. Value at Risk $V@R_\lambda$: $\Lambda(x)$: $\lambda \in (0, 1)$.

Take in the definition of the family (3.4), the function Λ to be equal to the constant λ, $\Lambda(x) = \lambda \in (0, 1)$. Then

$$F_m(x) := \lambda \mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x),$$
$$\mathcal{A}^m := \left\{ Q \in \mathcal{P} \mid F_Q \leq F_m \right\},$$
$$\Phi_{V@R_\lambda}(P) := -\sup\{m \in \mathbb{R} \mid P \in \mathcal{A}^m\}.$$

If the random variable $X \in L^0$ has distribution function P_X and $q_X^+(\lambda) = \sup\{x \in \mathbb{R} \mid \mathbb{P}(X \leq x) \leq \lambda\}$ is the right continuous inverse of P_X then

$$\begin{aligned}\Phi_{V@R_\lambda}(P_X) &= -\sup\{m \mid P_X \in \mathcal{A}^m\}\\ &= -\sup\{m \mid \mathbb{P}(X \leq x) \leq \lambda \ \forall x < m\}\\ &= -\sup\{m \mid \mathbb{P}(X \leq m) \leq \lambda\}\\ &= -q_X^+(\lambda) := V@R_\lambda(X)\end{aligned}$$

coincide with the Value at Risk of level $\lambda \in (0, 1)$. As the family $\{F_m\}$ is feasible, $\Phi_{V@R_\lambda} : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ is (Mon), (Qco), $\sigma(\mathcal{P}, C_b)$ -*lsc*. In addition, it also satisfies (TrI).

As well known, $V@R_\lambda$: $L^0 \to \mathbb{R} \cup \{\infty\}$ is not quasi-convex, as a map defined on random variables, even though the corresponding $\Phi_{V@R_\lambda} : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$, as a map defined on distribution functions, is quasi-convex (see Drapeau and Kupper 2010 for a discussion on this issue).

EXAMPLE 3.10. Fix the family $\{\Lambda_m\}_{m\in\mathbb{R}}$ of functions $\Lambda_m : \mathbb{R} \to [0, 1]$ such that for every $m \in \mathbb{R}$, $\Lambda_m(\cdot)$ is right continuous (w.r.t. x) and for every $x \in \mathbb{R}$, $\Lambda_\cdot(x)$ is decreasing and left continuous (w.r.t. m). Define the family $\{F_m\}_{m\in\mathbb{R}}$ of functions $F_m : \mathbb{R} \to [0, 1]$ by

$$F_m(x) := \Lambda_m(x)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x). \tag{3.5}$$

It is easy to check that if $\sup_{x\in\mathbb{R}} \Lambda_{m_0}(x) < 1$, for some $m_0 \in \mathbb{R}$, then the family $\{F_m\}_{m\in\mathbb{R}}$ is feasible and therefore the associated map $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ is well defined, (Mon), (Qco), $\sigma(\mathcal{P}, C_b)$-*lsc*.

4. ON THE $\Lambda V@R$

We now propose a generalization of the $V@R_\lambda$ which appears useful for possible application whenever an agent is facing some ambiguity on the parameter λ, namely λ is given by some uncertain value in a confidence interval $[\lambda^m, \lambda^M]$, with $0 \leq \lambda^m \leq \lambda^M \leq 1$. The $V@R_\lambda$ corresponds to case $\lambda^m = \lambda^M$ and one typical value is $\lambda^M = 0, 05$.

We will distinguish two possible classes of agents:

Risk Prudent Agents: Fix the *increasing* right continuous function $\Lambda : \mathbb{R} \to [0, 1]$, choose as in (3.4)

$$F_m(x) = \Lambda(x)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x)$$

and set $\lambda^m := \inf \Lambda \geq 0$, $\lambda^M := \sup \Lambda \leq 1$. As the function Λ is increasing, we are assigning to a lower loss a lower probability. In particular, given two possible choices Λ_1,

Λ_2 for two different agents, the condition $\Lambda_1 \leq \Lambda_2$ means that the agent 1 is more risk prudent than agent 2. Set, as in (3.1), $\mathcal{A}^m = \{Q \in \mathcal{P} \mid F_Q \leq F_m\}$ and define as in (3.2)

$$\Lambda V@R(P) := -\sup\{m \in \mathbb{R} \mid P \in \mathcal{A}^m\}.$$

Thus, in case of a random variable X

$$\Lambda V@R(P_X) := -\sup\{m \in \mathbb{R} \mid \mathbb{P}(X \leq x) \leq \Lambda(x),\ \forall x \leq m\}.$$

In particular, it can be rewritten as

$$\Lambda V@R(P_X) = -\inf\{x \in \mathbb{R} \mid \mathbb{P}(X \leq x) > \Lambda(x)\}.$$

If both F_X and Λ are continuous $\Lambda V@R$ corresponds to the smallest intersection between the two curves.

In this section, we assume that

$$\lambda^M < 1.$$

Besides its obvious financial motivation, this request implies that the corresponding family F_m is feasible and so $\Lambda V@R(P) > -\infty$ for all $P \in \mathcal{P}$.

The feasibility of the family $\{F_m\}$ implies that the $\Lambda V@R : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ is well defined, (Mon), (QCo), and (CfA) (or equivalently $\sigma(\mathcal{P}, C_b)$-*lsc*) map.

EXAMPLE 4.1. One possible simple choice of the function Λ is represented by the step function:

$$\Lambda(x) = \lambda^m \mathbf{1}_{(-\infty,\bar{x})}(x) + \lambda^M \mathbf{1}_{[\bar{x},+\infty)}(x).$$

The idea is that with a probability of λ^M we are accepting to loose at most $\bar{x}$. In this case we observe that:

$$\Lambda V@R(P) = \begin{cases} V@R_{\lambda^M}(P) & \text{if } V@R_{\lambda^m}(P) \leq -\bar{x} \\ V@R_{\lambda^m}(P) & \text{if } V@R_{\lambda^m}(P) > -\bar{x}. \end{cases}$$

Even though the $\Lambda V@R$ is continuous from above (Proposition 3.5 and 2.5), it may not be continuous from below, as this example shows. For instance, take $\bar{x} = 0$ and P_{X_n} induced by a sequence of uniformly distributed random variables $X_n \sim U\left[-\lambda^m - \frac{1}{n}, 1 - \lambda^m - \frac{1}{n}\right]$. We have $P_{X_n} \uparrow P_{U[-\lambda^m, 1-\lambda^m]}$ but $\Lambda V@R(P_{X_n}) = -\frac{1}{n}$ for every n and $\Lambda V@R(P_{U[-\lambda^m,1-\lambda^m]}) = \lambda^M - \lambda^m$.

REMARK 4.2.

(i) If $\lambda^m = 0$ the domain of $\Lambda V@R(P)$ is not the entire convex set $\mathcal{P}$. We have two possible cases
- $\mathrm{supp}(\Lambda) = [x^*, +\infty)$: in this case $\Lambda V@R(P) = -\inf \mathrm{supp}(F_P)$ for every $P \in \mathcal{P}$ such that $\mathrm{supp}(F_P) \supseteq \mathrm{supp}(\Lambda)$.
- $\mathrm{supp}(\Lambda) = (-\infty, +\infty)$: in this case

$$\Lambda V@R(P) = +\infty \quad \text{for all } P, \text{ such that } \lim_{x \to -\infty} \frac{F_P(x)}{\Lambda(x)} > 1,$$
$$\Lambda V@R(P) < +\infty \quad \text{for all } P, \text{ such that } \lim_{x \to -\infty} \frac{F_P(x)}{\Lambda(x)} < 1.$$

In the case $\lim_{x\to-\infty} \frac{F_P(x)}{\Lambda(x)} = 1$ both the previous behaviors might occur.

(ii) In case that $\lambda^m > 0$ then $\Lambda V@R(P) < +\infty$ for all $P \in \mathcal{P}$, so that $\Lambda V@R$ is finite valued.

We can prove a further structural property which is the counterpart of (TrI) for the $\Lambda V@R$. Let $\alpha \in \mathbb{R}$ any cash amount

$$\begin{aligned}\Lambda V@R(P_{X+\alpha}) &= -\sup\{m \mid \mathbb{P}(X+\alpha \le x) \le \Lambda(x),\ \forall x \le m\}\\ &= -\sup\{m \mid \mathbb{P}(X \le x-\alpha) \le \Lambda(x),\ \forall x \le m\}\\ &= -\sup\{m \mid \mathbb{P}(X \le y) \le \Lambda(y+\alpha),\ \forall y \le m-\alpha\}\\ &= -\sup\{m+\alpha \mid \mathbb{P}(X \le y) \le \Lambda(y+\alpha),\ \forall y \le m\}\\ &= \Lambda^\alpha V@R(P_X) - \alpha,\end{aligned}$$

where $\Lambda^\alpha(x) = \Lambda(x+\alpha)$. We may conclude that if we add a sure positive (resp. negative) amount α to a risky position X then the risk decreases (resp. increases) of the value $-\alpha$, constrained to a lower (resp. higher) level of risk prudence described by $\Lambda^\alpha \ge \Lambda$ (resp. $\Lambda^\alpha \le \Lambda$). For an arbitrary $P \in \mathcal{P}$ this property can be written as

$$\Lambda V@R(T_\alpha P) = \Lambda^\alpha V@R(P) - \alpha, \quad \forall \alpha \in \mathbb{R},$$

where $T_\alpha P(-\infty, x] = P(-\infty, x-\alpha]$.

Risk Seeking Agents: Fix the *decreasing* right continuous function $\Lambda : \mathbb{R} \to [0,1]$, with $\inf \Lambda < 1$. Similarly as above, we define

$$F_m(x) = \Lambda(x)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x)$$

and the (Mon), (QCo), and (CfA) map

$$\Lambda V@R(P) := -\sup\{m \in \mathbb{R} \mid F_P \le F_m\} = -\sup\{m \in \mathbb{R} \mid \mathbb{P}(X \le m) \le \Lambda(m)\}.$$

In this case, for eventual huge losses we are allowing the highest level of probability. As in the previous example let $\alpha \in \mathbb{R}$ and note that

$$\Lambda V@R(P_{X+\alpha}) = \Lambda^\alpha V@R(P_X) - \alpha,$$

where $\Lambda^\alpha(x) = \Lambda(x+\alpha)$. The property is exactly the same as in the former example but here the interpretation is slightly different. If we add a sure positive (resp. negative) amount α to a risky position X then the risk decreases (resp. increases) of the value $-\alpha$, constrained to a lower (resp. higher) level of risk seeking because $\Lambda^\alpha \le \Lambda$ (resp. $\Lambda^\alpha \ge \Lambda$).

REMARK 4.3. For a decreasing Λ, there is a simpler formulation—which will be used in Section 5.3—of the $\Lambda V@R$ that is obtained replacing in F_m the function Λ with the line $\Lambda(m)$ for all $x < m$. Let

$$\tilde{F}_m(x) = \Lambda(m)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x).$$

This family is of the type (3.5) and is feasible, provided the function Λ is continuous. For a decreasing Λ, it is evident that

$$\Lambda V@R(P) = \Lambda \widetilde{V}@R(P) := -\sup\left\{m \in \mathbb{R} \mid F_P \le \tilde{F}_m\right\},$$

as the function Λ lies above the line $\Lambda(m)$ for all $x \le m$.

5. QUASI-CONVEX DUALITY

In literature we also find several results about the dual representation of law invariant risk measures. Kusuoka (2001) contributed to the coherent case, although Frittelli and Rosazza Gianin (2005) extended this result to the convex case. Jouini, Schachermayer, and Touzi (2006) and Filipovic and Svindland (2012), in the convex case, and Svindland (2010) in the quasi-convex case, showed that every law invariant risk measure is already weakly lower semicontinuous. Recently, Cerreia-Vioglio et al. (2011b) provided a robust dual representation for law invariant quasi-convex risk measures, which has been extended to the dynamic case by Frittelli and Maggis (2011a, 2011b).

In Sections 5.1 and 5.2 we will treat the general case of maps defined on $\mathcal{P}$, although in Section 5.3 we specialize these results to show the dual representation of maps associated to feasible families.

5.1. Reasons of the Failure of the Convex Duality for Translation Invariant Maps on $\mathcal{P}$

It is well known that the classical convex duality provided by the Fenchel-Moreau theorem (Fenchel 1949) guarantees the representation of convex and lower semicontinuous functions and therefore is very useful for the dual representation of convex risk measures (see Frittelli and Rosazza Gianin 2002). For any map $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ let Φ^* be the convex conjugate:

$$\Phi^*(f) := \sup_{Q \in \mathcal{P}} \left\{ \int f dQ - \Phi(Q) \right\}, \ f \in C_b.$$

Applying the fact that $\mathcal{P}$ is a $\sigma(ca, C_b)$-closed convex subset of ca one can easily check that the following version of Fenchel-Moreau Theorem holds true for maps defined on $\mathcal{P}$.

PROPOSITION 5.1 [Fenchel-Moreau]. *Suppose that* $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ *is* $\sigma(\mathcal{P}, C_b)$*-lsc and convex. If* $Dom(\Phi) := \{Q \in \mathcal{P} \mid \Phi(Q) < +\infty\} \neq \varnothing$ *then* $Dom(\Phi^*) \neq \varnothing$ *and*

$$\Phi(Q) = \sup_{f \in C_b} \left\{ \int f dQ - \Phi^*(f) \right\}.$$

One trivial example of a proper $\sigma(\mathcal{P}, C_b)$-*lsc* and convex map on $\mathcal{P}$ is given by $Q \to \int f dQ$, for some $f \in C_b$. But this map does not satisfy the (TrI) property. Indeed, we show that in the setting of risk measures defined on $\mathcal{P}$, weakly lower semicontinuity and convexity are incompatible with translation invariance.

PROPOSITION 5.2. *For any map* $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$*, if there exists a sequence* $\{Q_n\}_n \subseteq \mathcal{P}$ *such that* $\underline{\lim}_n \Phi(Q_n) = -\infty$ then $Dom(\Phi^*) = \varnothing$.

Proof. For any $f \in C_b(\mathbb{R})$

$$\Phi^*(f) = \sup_{Q \in \mathcal{P}} \left\{ \int f dQ - \Phi(Q) \right\} \geq \int f d(Q_n) - \Phi(Q_n) \geq \inf_{x \in \mathbb{R}} f(x) - \Phi(Q_n),$$

which implies $\Phi^* = +\infty$. □

From Propositions (5.1) and (5.2) we immediately obtain:

COROLLARY 5.3. *Let $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ be $\sigma(\mathcal{P}, C_b)$-lsc, convex and not identically equal to $+\infty$. Then Φ is not (TrI), is not cash sup additive (i.e., it does not satisfy: $\Phi(T_m Q) \leq \Phi(Q) - m$) and $\underline{\lim}_n \Phi(\delta_n) \neq -\infty$. In particular, the certainty equivalent maps Φ_f defined in (2.1) can not be convex, as they are $\sigma(\mathcal{P}, C_b)$-lsc and $\Phi_f(\delta_n) = -n$.*

5.2. The Dual Representation

As described in the examples in Section 3, the $\Phi_{V@R_\lambda}$ and Φ_w are proper, $\sigma(ca, C_b)$-*lsc*, quasi-convex, (Mon), and (TrI) maps $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$. Therefore, the negative result outlined in Corollary 5.3 for the convex case can not be true in the quasi-convex setting.

We recall that the seminal contribution to quasi-convex duality comes from the dual representation by Volle (1998) and Penot and Volle (1990), which has been sharpened to a complete quasi-convex duality by Cerreia-Vioglio et al. (2011b) (case of M-spaces), Cerreia-Vioglio (2009) (preferences over menus), and Drapeau and Kupper (2010) (for general topological vector spaces).

Here we replicate this result and provide the dual representation of a $\sigma(\mathcal{P}, C_b)$-*lsc* quasi-convex maps defined on the entire set $\mathcal{P}$. The main difference is that our map Φ is defined on a convex subset of *ca* and not a vector space (a similar result can be found in Drapeau and Kupper 2010 for convex sets). But because $\mathcal{P}$ is $\sigma(ca, C_b)$-closed, the first part of the proof will match very closely the one given by Volle. To achieve the dual representation of $\sigma(\mathcal{P}, C_b)$-*lsc* risk measures $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ we will impose the monotonicity assumption of Φ and deduce that in the dual representation the supremum can be restricted to the set

$$C_b^- = \{ f \in C_b \mid f \text{ is decreasing} \} .$$

This is natural as the first-order stochastic dominance implies (see theorem 2.70 in Föllmer and Schied 2004) that

$$C_b^- = \left\{ f \in C_b \mid Q, P \in \mathcal{P} \text{ and } P \preccurlyeq Q \Rightarrow \int f dQ \leq \int f dP \right\} . \tag{5.1}$$

Note that differently from Drapeau and Kupper (2010) the following proposition does not require the extension of the risk map to the entire space $ca(\mathbb{R})$. Once the representation is obtained the uniqueness of the dual function is a direct consequence of theorem 2.19 in Drapeau and Kupper (2010) as explained by Proposition 5.9.

PROPOSITION 5.4.

(i) *Any $\sigma(\mathcal{P}, C_b)$-lsc and quasi-convex functional $\Phi : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ can be represented as*

$$\Phi(P) = \sup_{f \in C_b} R\left(\int f dP, f \right), \tag{5.2}$$

where $R : \mathbb{R} \times C_b \to \overline{\mathbb{R}}$ is defined by

$$R(t, f) := \inf_{Q \in \mathcal{P}} \left\{ \Phi(Q) \mid \int f dQ \geq t \right\} . \tag{5.3}$$

(ii) *If in addition Φ is monotone then (5.2) holds with C_b replaced by C_b^-.*

Proof. We will use the fact that $\sigma(\mathcal{P}, C_b)$ is the relativization of $\sigma(ca, C_b)$ to the set $\mathcal{P}$. In particular, the lower level sets will be $\sigma(ca, C_b)$-closed.

(i) By definition, for any $f \in C_b(\mathbb{R})$, $R(\int f dP, f) \leq \Phi(P)$ and therefore

$$\sup_{f \in C_b} R\left(\int f dP, f\right) \leq \Phi(P), \quad P \in \mathcal{P}.$$

Fix any $P \in \mathcal{P}$ and take $\varepsilon \in \mathbb{R}$ such that $\varepsilon > 0$. Then P does not belong to the $\sigma(ca, C_b)$-closed convex set

$$\mathcal{C}_\varepsilon := \{Q \in \mathcal{P} : \Phi(Q) \leq \Phi(P) - \varepsilon\}$$

(if $\Phi(P) = +\infty$, replace the set $\mathcal{C}_\varepsilon$ with $\{Q \in \mathcal{P} : \Phi(Q) \leq M\}$, for any M). By the Hahn Banach theorem there exists a continuous linear functional that strongly separates P and $\mathcal{C}_\varepsilon$, i.e., there exists $\alpha \in \mathbb{R}$ and $f_\varepsilon \in C_b$ such that

$$\int f_\varepsilon dP > \alpha > \int f_\varepsilon dQ \quad \text{for all } Q \in \mathcal{C}_\varepsilon. \tag{5.4}$$

Hence:

$$\left\{Q \in \mathcal{P} : \int f_\varepsilon dP \leq \int f_\varepsilon dQ\right\} \subseteq (\mathcal{C}_\varepsilon)^C = \{Q \in \mathcal{P} : \Phi(Q) > \Phi(P) - \varepsilon\} \tag{5.5}$$

and

$$\begin{aligned} \Phi(P) &\geq \sup_{f \in C_b} R\left(\int f dP, f\right) \geq R\left(\textstyle\int f_\varepsilon dP, f_\varepsilon\right) \\ &= \inf\left\{\Phi(Q) \mid Q \in \mathcal{P} \text{ such that } \int f_\varepsilon dP \leq \int f_\varepsilon dQ\right\} \\ &\geq \inf\{\Phi(Q) \mid Q \in \mathcal{P} \text{ satisfying } \Phi(Q) > \Phi(P) - \varepsilon\} \geq \Phi(P) - \varepsilon. \end{aligned} \tag{5.6}$$

(ii) We furthermore assume that Φ is monotone. As shown in (i), for every $\varepsilon > 0$ we find f_ε such that (5.4) holds true. We claim that there exists $g_\varepsilon \in C_b^-$ satisfying:

$$\int g_\varepsilon dP > \alpha > \int g_\varepsilon dQ \quad \text{for all } Q \in \mathcal{C}_\varepsilon, \tag{5.7}$$

and then the above argument (in equations (5.4)–(5.6)) implies the thesis.

We define the decreasing function

$$g_\varepsilon(x) =: \sup_{y \geq x} f_\varepsilon(y) \in C_b^-.$$

First case: Suppose that $g_\varepsilon(x) = \sup_{x \in \mathbb{R}} f_\varepsilon(x) =: s$. In this case there exists a sequence of $\{x_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}$ such that $x_n \to +\infty$ and $f_\varepsilon(x_n) \to s$, as $n \to \infty$. Define

$$g_n(x) = s\mathbf{1}_{(-\infty, x_n]} + f_\varepsilon(x)\mathbf{1}_{(x_n, +\infty)}$$

and note that $s \geq g_n \geq f_\varepsilon$ and $g_n \uparrow s$. For any $Q \in \mathcal{C}_\varepsilon$ we consider Q_n defined by $F_{Q_n}(x) = F_Q(x)\mathbf{1}_{[x_n,+\infty)}$. Because $Q \preccurlyeq Q_n$, monotonicity of Φ implies $Q_n \in \mathcal{C}_\varepsilon$. Note that

$$\int g_n dQ - \int f_\varepsilon dQ_n = (s - f_\varepsilon(x_n))Q(-\infty, x_n] \overset{n\to+\infty}{\longrightarrow} 0, \text{ as } n \to \infty. \tag{5.8}$$

From equation (5.4) we have

$$s \geq \int f_\varepsilon dP > \alpha > \int f_\varepsilon dQ_n \quad \text{for all } n \in \mathbb{N}. \tag{5.9}$$

Letting $\delta = s - \alpha > 0$ we obtain $s > \int f_\varepsilon dQ_n + \frac{\delta}{2}$. From (5.8), there exists $\overline{n} \in \mathbb{N}$ such that $0 \leq \int g_n dQ - \int f_\varepsilon dQ_n < \frac{\delta}{4}$ for every $n \geq \overline{n}$. Therefore, $\forall\, n \geq \overline{n}$

$$s > \int f_\varepsilon dQ_n + \frac{\delta}{2} > \int g_n dQ - \frac{\delta}{4} + \frac{\delta}{2} = \int g_n dQ + \frac{\delta}{4}$$

and this leads to a contradiction as $g_n \uparrow s$. So the first case is excluded.

Second case: Suppose that $g_\varepsilon(x) < s$ for any $x > \overline{x}$. As the function $g_\varepsilon \in C_b^-$ is decreasing, there will exists at most a countable sequence of intervals $\{A_n\}_{n\geq 0}$ on which g_ε is constant. Set $A_0 = (-\infty, b_0)$, $A_n = [a_n, b_n) \subset \mathbb{R}$ for $n \geq 1$. W.l.o.g. we suppose that $A_n \cap A_m = \emptyset$ for all $n \neq m$ (else, we paste together the sets) and $a_n < a_{n+1}$ for every $n \geq 1$. We stress that $f_\varepsilon(x) = g_\varepsilon(x)$ on $D =: \bigcap_{n\geq 0} A_n^C$. For every $Q \in \mathcal{C}_\varepsilon$ we define the probability $\overline{Q}$ by its distribution function as

$$F_{\overline{Q}}(x) = F_Q(x)\mathbf{1}_D + \sum_{n\geq 1} F_Q(a_n)\mathbf{1}_{[a_n,b_n)}.$$

As before, $Q \preccurlyeq \overline{Q}$ and monotonicity of Φ implies $\overline{Q} \in \mathcal{C}_\varepsilon$. Moreover,

$$\int g_\varepsilon dQ = \int_D f_\varepsilon dQ + f_\varepsilon(b_0)Q(A_0) + \sum_{n\geq 1} f_\varepsilon(a_n)Q(A_n) = \int f_\varepsilon d\overline{Q}.$$

From $g_\varepsilon \geq f_\varepsilon$ and equation (5.4) we deduce

$$\int g_\varepsilon dP \geq \int f_\varepsilon dP > \alpha > \int f_\varepsilon d\overline{Q} = \int g_\varepsilon dQ \quad \text{for all } Q \in \mathcal{C}_\varepsilon.$$

□

We reformulate the Proposition 5.4 and provide two dual representation of $\sigma(\mathcal{P}(\mathbb{R}), C_b)$-*lsc* Risk Measure $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ in terms of a supremum over a class of probabilistic scenarios. Let

$$\mathcal{P}_c(\mathbb{R}) = \left\{Q \in \mathcal{P}(\mathbb{R}) \mid F_Q \text{ is continuous}\right\}.$$

PROPOSITION 5.5. *Any $\sigma(\mathcal{P}(\mathbb{R}), C_b)$-lsc Risk Measure $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ can be represented as*

$$\Phi(P) = \sup_{Q\in\mathcal{P}_c(\mathbb{R})} R\left(-\int F_Q dP, -F_Q\right).$$

Proof. Note that for every $f \in C_b^-$ which is constant we have $R(\int f dP, f) = \inf_{Q \in \mathcal{P}} \Phi(Q)$. Therefore, we may assume w.l.o.g. that $f \in C_b^-$ is not constant. Then $g := \frac{f - f(+\infty)}{f(-\infty) - f(+\infty)} \in C_b^-$, $\inf g = 0$, $\sup g = 1$, and so: $g \in \{-F_Q \mid Q \in \mathcal{P}_c(\mathbb{R})\}$. In addition, because $\int f dQ \geq \int f dP$ i.f.f. $\int g dQ \geq \int g dP$ we obtain from (5.2) and (ii) of Proposition 5.4

$$\Phi(P) = \sup_{f \in C_b^-} R\left(\int f dP, f\right) = \sup_{Q \in \mathcal{P}_c(\mathbb{R})} R\left(-\int F_Q dP, -F_Q\right).$$

□

Finally, we state the dual representations for Risk Measures expressed either in terms of the dual function R as used by Cerreia-Vioglio et al. (2011b), or considering the left continuous version of R (see Lemma 5.7) in the formulation proposed by Drapeau and Kupper (2010). If $R : \mathbb{R} \times C_b(\mathbb{R}) \to \overline{\mathbb{R}}$, the left continuous version of $R(\cdot, f)$ is defined by:

$$R^-(t, f) := \sup\{R(s, f) \mid s < t\}. \tag{5.10}$$

PROPOSITION 5.6. *Any $\sigma(\mathcal{P}(\mathbb{R}), C_b)$-lsc Risk Measure $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ can be represented as*

$$\Phi(P) = \sup_{f \in C_b^-} R\left(\int f dP, f\right) = \sup_{f \in C_b^-} R^-\left(\int f dP, f\right). \tag{5.11}$$

The function $R^-(t, f)$ defined in (5.10) *can be written as*

$$R^-(t, f) = \inf\{m \in \mathbb{R} \mid \gamma(m, f) \geq t\}, \tag{5.12}$$

where $\gamma : \mathbb{R} \times C_b(\mathbb{R}) \to \overline{\mathbb{R}}$ is given by:

$$\gamma(m, f) := \sup_{Q \in \mathcal{P}} \left\{\int f dQ \mid \Phi(Q) \leq m\right\}, m \in \mathbb{R}. \tag{5.13}$$

Proof. Note that $R(\cdot, f)$ is increasing and $R(t, f) \geq R^-(t, f)$. If $f \in C_b^-$ then $P \preccurlyeq Q \Rightarrow \int f dQ \leq \int f dP$. Therefore,

$$R^-\left(\int f dP, f\right) := \sup_{s < \int f dP} R(s, f) \geq \lim_{P_n \downarrow P} R\left(\int f dP_n, f\right).$$

From Proposition 5.4 (ii) we obtain:

$$\begin{aligned}\Phi(P) &= \sup_{f \in C_b^-} R\left(\int f dP, f\right) \geq \sup_{f \in C_b^-} R^-\left(\int f dP, f\right) \geq \sup_{f \in C_b^-} \lim_{P_n \downarrow P} R\left(\int f dP_n, f\right) \\ &= \lim_{P_n \downarrow P} \sup_{f \in C_b^-} R\left(\int f dP_n, f\right) = \lim_{P_n \downarrow P} \Phi(P_n) = \Phi(P)\end{aligned}$$

by (CfA). This proves (5.11). The second statement follows from Lemma 5.7. □

The following lemma shows that the left continuous version of R is the left inverse of the function γ as defined in 5.13 (for the definition and the properties of the left inverse we refer to Föllmer and Schied 2004, section A.3).

LEMMA 5.7. *Let Φ be any map $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ and $R : \mathbb{R} \times C_b(\mathbb{R}) \to \overline{\mathbb{R}}$ be defined in (5.3). The left continuous version of $R(\cdot, f)$ can be written as:*

$$R^{-}(t, f) := \sup\{R(s, f) \mid s < t\} = \inf\{m \in \mathbb{R} \mid \gamma(m, f) \geq t\}, \tag{5.14}$$

where $\gamma : \mathbb{R} \times C_b(\mathbb{R}) \to \overline{\mathbb{R}}$ is given in (5.13).

Proof. Let the right-hand side (RHS) of equation (5.14) be denoted by

$$S(t, f) := \inf\{m \in \mathbb{R} \mid \gamma(m, f) \geq t\}, \ (t, f) \in \mathbb{R} \times C_b(\mathbb{R}),$$

and note that $S(\cdot, f)$ is the left inverse of the increasing function $\gamma(\cdot, f)$ and therefore $S(\cdot, f)$ is left continuous.
Step I: To prove that $R^{-}(t, f) \geq S(t, f)$ it is sufficient to show that for all $s < t$ we have:

$$R(s, f) \geq S(s, f). \tag{5.15}$$

Indeed, if (5.15) is true

$$R^{-}(t, f) = \sup_{s<t} R(s, f) \geq \sup_{s<t} S(s, f) = S(t, f),$$

as both R^{-} and S are left continuous in the first argument. Writing explicitly the inequality (5.15)

$$\inf_{Q \in \mathcal{P}} \left\{\Phi(Q) \mid \int f\,dQ \geq s\right\} \geq \inf\{m \in \mathbb{R} \mid \gamma(m, f) \geq s\}$$

and letting $Q \in \mathcal{P}$ satisfying $\int f dQ \geq s$, we see that it is sufficient to show the existence of $m \in \mathbb{R}$ such that $\gamma(m, f) \geq s$ and $m \leq \Phi(Q)$. If $\Phi(Q) = -\infty$ then $\gamma(m, f) \geq s$ for any m and therefore $S(s, f) = R(s, f) = -\infty$.

Suppose now that $+\infty > \Phi(Q) > -\infty$ and define $m := \Phi(Q)$. As $\int f dQ \geq s$ we have:

$$\gamma(m, f) := \sup_{Q \in \mathcal{P}} \left\{\int f\,dQ \mid \Phi(Q) \leq m\right\} \geq s.$$

Then $m \in \mathbb{R}$ satisfies the required conditions.

Step II: To obtain $R^{-}(t, f) := \sup_{s<t} R(s, f) \leq S(t, f)$ it is sufficient to prove that, for all $s < t$, $R(s, f) \leq S(t, f)$, that is

$$\inf_{Q \in \mathcal{P}} \left\{\Phi(Q) \mid \int f\,dQ \geq s\right\} \leq \inf\{m \in \mathbb{R} \mid \gamma(m, f) \geq t\}. \tag{5.16}$$

Fix any $s < t$ and consider any $m \in \mathbb{R}$ such that $\gamma(m, f) \geq t$. By the definition of γ, for all $\varepsilon > 0$ there exists $Q_\varepsilon \in \mathcal{P}$ such that $\Phi(Q_\varepsilon) \leq m$ and $\int f dQ_\varepsilon > t - \varepsilon$. Take ε such that $0 < \varepsilon < t - s$. Then $\int f dQ_\varepsilon \geq s$ and $\Phi(Q_\varepsilon) \leq m$ and (5.16) follows. $\square$

Complete duality: The complete duality in the class of quasi-convex monotone maps on vector spaces was first obtained by Cerreia-Vioglio et al. (2011a). The following proposition is based on the complete duality proved in Drapeau and Kupper (2010) for maps defined on convex sets and therefore the results in Drapeau and Kupper (2010) apply very easily in our setting. To obtain the uniqueness of the dual function in the representation (5.11) we need to introduce the opportune class $\mathcal{R}^{\max}$. Recall that $\mathcal{P}(\mathbb{R})$ spans the space of countably additive signed measures on $\mathbb{R}$, namely $ca(\mathbb{R})$ and that the first stochastic order corresponds to the cone

$$\mathcal{K} = \left\{ \mu \in ca \mid \int f d\mu \geq 0 \ \forall\, f \in \mathcal{K}^{\circ} \right\} \subseteq ca_{+},$$

where $\mathcal{K}^{\circ} = -C_b^{-}$ are the nondecreasing functions $f \in C_b$.

Definition 5.8 (Drapeau and Kupper 2010). We denote by $\mathcal{R}^{\max}$ the class of functions $R : \mathbb{R} \times \mathcal{K}^{\circ} \to \overline{\mathbb{R}}$ such that: (i) R is nondecreasing and left continuous in the first argument,(ii) R is jointly quasi-concave, (iii) $R(s, \lambda f) = R(\frac{s}{\lambda}, f)$ for every $f \in \mathcal{K}^{\circ}$, $s \in \mathbb{R}$ and $\lambda > 0$, (iv) $\lim_{s \to -\infty} R(s, f) = \lim_{s \to -\infty} R(s, g)$ for every $f, g \in \mathcal{K}^{\circ}$, (v) $R^{+}(s, f) = \inf_{s' > s} R(s', f)$, is upper semicontinuous in the second argument.

Proposition 5.9. *Any $\sigma(\mathcal{P}(\mathbb{R}), C_b)$-lsc Risk Measure $\Phi : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ can be represented as in 5.11. The function $R^{-}(t, f)$ given by 5.12 is unique in the class $\mathcal{R}^{\max}$.*

Proof. According to Definition 2.13 in Drapeau and Kupper (2010) a map $\Phi : \mathcal{P} \to \overline{R}$ is continuously extensible to ca if

$$\overline{\mathcal{A}^{m} + \mathcal{K}} \cap \mathcal{P} = \mathcal{A}^{m},$$

where $\mathcal{A}^{m}$ is acceptance set of level m and $\mathcal{K}$ is the ordering positive cone on ca. Observe that $\mu \in ca_{+}$ satisfies $\mu(E) \geq 0$ for every $E \in \mathcal{B}_{\mathbb{R}}$ so that $P + \mu \notin \mathcal{P}$ for $P \in \mathcal{A}^{m}$ and $\mu \in \mathcal{K}$ except if $\mu = 0$. For this reason the *lsc* map Φ admits a lower semicontinuous extension to ca and then theorem 2.19 in Drapeau and Kupper (2010) applies and we get the uniqueness in the class $\mathcal{R}^{\max}_{\mathcal{P}}$ (see definition 2.17 in Drapeau and Kupper 2010). In addition, $\mathcal{R}^{\max} = \mathcal{R}^{\max}_{\mathcal{P}}$ follows exactly by the same argument at the end of the proof of proposition 3.5 (Drapeau and Kupper 2010). Finally, we note that lemma C.2 in Drapeau and Kupper (2010) implies that $R^{-} \in \mathcal{R}^{\max}$ as $\gamma(m, f)$ is convex, positively homogeneous, and *lsc* in the second argument. $\square$

5.3. Computation of the Dual Function

The following proposition is useful to compute the dual function $R^{-}(t, f)$ for the examples considered in this paper.

Proposition 5.10. *Let $\{F_m\}_{m \in \mathbb{R}}$ be a feasible family and suppose in addition that, for every m, $F_m(x)$ is increasing in x and $\lim_{x \to +\infty} F_m(x) = 1$. The associated map $\Phi : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ defined in (3.2) is well defined, (Mon), (Qco), and $\sigma(\mathcal{P}, C_b)$-lsc and the representation (5.11) holds true with R^{-} given in (5.12) and*

$$\gamma(m, f) = \int f dF_{-m} + F_{-m}(-\infty) f(-\infty). \tag{5.17}$$

Proof. From equations (3.1) and (3.3) we obtain:

$$\mathcal{A}^{-m} = \left\{Q \in \mathcal{P}(\mathbb{R}) \mid F_Q \leq F_{-m}\right\} = \{Q \in \mathcal{P} \mid \Phi(Q) \leq m\}$$

so that

$$\gamma(m, f) := \sup_{Q\in\mathcal{P}} \left\{\int f dQ \mid \Phi(Q) \leq m\right\} = \sup_{Q\in\mathcal{P}} \left\{\int f dQ \mid F_Q \leq F_{-m}\right\}.$$

Fix $m \in \mathbb{R}$, $f \in C_b^-$ and define the distribution function $F_{Q_n}(x) = F_{-m}(x)\mathbf{1}_{[-n,+\infty)}$ for every $n \in \mathbb{N}$. Obviously, $F_{Q_n} \leq F_{-m}$, $Q_n \downarrow$ and, taking into account (5.1), $\int fdQ_n$ is increasing. For any $\varepsilon > 0$, let $Q^\varepsilon \in \mathcal{P}$ satisfy $F_{Q^\varepsilon} \leq F_{-m}$ and $\int fdQ^\varepsilon > \gamma(m,f) - \varepsilon$. Then: $F_{Q_n^\varepsilon}(x) := F_{Q^\varepsilon}(x)\mathbf{1}_{[-n,+\infty)} \uparrow F_{Q^\varepsilon}$, $F_{Q_n^\varepsilon} \leq F_{Q_n}$ and

$$\int f dQ_n \geq \int f dQ_n^\varepsilon \uparrow \int f dQ^\varepsilon > \gamma(m, f) - \varepsilon.$$

We deduce that $\int fdQ_n \uparrow \gamma(m,f)$ and, because

$$\int f dQ_n = \int_{-n}^{+\infty} f dF_{-m} + F_{-m}(-n)f(-n),$$

we obtain (5.17). □

EXAMPLE 5.11. Computation of $\gamma(m,f)$ for the $\Lambda V@R$.

Let $m \in \mathbb{R}$ and $f \in C_b^-$. As $F_m(x) = \Lambda(x)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x)$, we compute from (5.17):

$$\gamma(m, f) = \int_{-\infty}^{-m} f d\Lambda + (1 - \Lambda(-m))f(-m) + \Lambda(-\infty)f(-\infty). \tag{5.18}$$

We apply the integration by parts and deduce

$$\int_{-\infty}^{-m} \Lambda df = \Lambda(-m)f(-m) - \Lambda(-\infty)f(-\infty) - \int_{-\infty}^{-m} f d\Lambda.$$

We can now substitute in equation (5.18) and get:

$$\gamma(m, f) = f(-m) - \int_{-\infty}^{-m} \Lambda df = f(-\infty) + \int_{-\infty}^{-m} (1 - \Lambda)df, \tag{5.19}$$

$$R^-(t, f) = -H_f^l(t - f(-\infty)), \tag{5.20}$$

where H_f^l is the left inverse of the function: $m \to \int_{-\infty}^{m}(1 - \Lambda)df$.

As a particular case, we match the results obtained in Drapeau and Kupper (2010) for the $V@R$ and the worst case risk measure. Indeed, from (5.19) and (5.20) we get: $R^-(t, f) = -f^l(\frac{t-\lambda f(-\infty)}{1-\lambda})$, if $\Lambda(x) = \lambda$; $R^-(t,f) = -f^l(t)$, if $\Lambda(x) = 0$, where f^l is the left inverse of f.

If Λ is decreasing we may use Remark 4.3 to derive a simpler formula for γ. Indeed, $\Lambda V@R(P) = \Lambda \widetilde{V}@R(P)$, where $\forall m \in \mathbb{R}$,

$$\widetilde{F}_m(x) = \Lambda(m)\mathbf{1}_{(-\infty,m)}(x) + \mathbf{1}_{[m,+\infty)}(x)$$

and so from (5.19)

$$\gamma(m, f) = f(-\infty) + [1 - \Lambda(-m)] \int_{-\infty}^{-m} df = [1 - \Lambda(-m)] f(-m) + \Lambda(-m) f(-\infty).$$

References

ALIPRANTIS, C. D., and K. C. BORDER (2005): *Infinite Dimensional Analysis*, 3rd ed., Berlin: Springer.

ARTZNER, P., F. DELBAEN, J. M. EBER, and D. HEATH (1999): Coherent Measures of Risk, *Math. Finance* 4, 203–228.

BIAGINI, S., and M. FRITTELLI (2009): On the Extension of the Namioka-Klee Theorem and on the Fatou Property for Risk Measures, in: *Optimality and Risk: Modern Trends in Mathematical Finance*, F. DELBAEN, M. RASONYI, and Ch. STRICKER, eds., Berlin: Springer, pp. 1–29.

CERREIA-VIOGLIO, S. (2009): Maxmin Expected Utility on a Subjective State Space: Convex Preferences under Risk, Preprint.

CERREIA-VIOGLIO, S., F. MACCHERONI, M. MARINACCI, and L. MONTRUCCHIO (2011a): Complete Monotone Quasiconcave Duality, *Math. Op. Res*. 36(2), 321–339.

CERREIA-VIOGLIO, S., F. MACCHERONI, M. MARINACCI, and L. MONTRUCCHIO (2011b): Risk Measures: Rationality and Diversification, *Math. Finance* 21, 743–774.

CHERNY, A., and D. MADAN (2009): New Measures for Performance Evaluation, *Rev. Financ. Stud.* 22, 2571–2606.

DRAPEAU, S., and M. KUPPER (2010): Risk Preferences and their Robust Representation. Forthcoming in *Math. Op. Res*.

FENCHEL, W. (1949): On Conjugate Convex Functions, *Can. J. Math.* 1, 73–77.

FILIPOVIC, D., and G. SVINDLAND (2012): The Canonical Model Space for Law Invariant Convex Risk Measures Is L, *Math. Finance* 22, 585–589.

FÖLLMER, H., and A. SCHIED (2004): *Stochastic Finance. An Introduction in Discrete Time*, 2nd ed., de Gruyter Studies in Mathematics, 27, Berlin: Walter de Gruyter & Co.

FRITTELLI, M., and M. MAGGIS (2011a): Dual Representation of Quasiconvex Conditional Maps, *SIAM J. Fin. Math.* 2, 357–382.

FRITTELLI, M., and M. MAGGIS (2011b): Conditional Certainty Equivalent, *Int. J. Theor. Appl. Finance* 14(1), 41–59.

FRITTELLI, M., and E. Rosazza GIANIN (2002): Putting Order in Risk Measures, *J. Banking Finance* 26(7), 1473–1486.

FRITTELLI, M., and E. Rosazza GIANIN (2005): Law Invariant Convex Risk Measures, *Adv. Math. Econ.* 7, 33–46.

JOUINI, E., W. SCHACHERMAYER, and N. TOUZI (2006): Law Invariant Risk Measures Have the Fatou Property, *Adv. Math. Econ.* 9, 49–71.

KUSUOKA, S. (2001): On Law Invariant Coherent Risk Measures, *Adv. Math. Econ.* 3, 83–95.

PENOT, J. P., and M. VOLLE (1990): On Quasi-Convex Duality, *Math. Oper. Res.* 15(1990), 597–625.

SHIRYAEV, A. N. (1995): *Probability*, *Graduate Texts in Mathematics*, 2nd ed., New York: Springer-Verlag.

SVINDLAND, G. (2010): Continuity Properties of Law Invariant (Quasi-) Convex Risk Functions, *Math. Financ. Econ.* 3(1), 39–43.

VOLLE, M. (1998): Duality for the Level Sum of Quasiconvex Functions and Applications, *ESAIM: Control, Optimisation and Calculus of Variations* 3, 329–343.

WEBER, S. (2006): Distribution Invariant Risk Measures, Information, and Dynamic Consistency, *Math. Finance* 16(2), 419–441.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 464–484

CONVEX RISK MEASURES FOR GOOD DEAL BOUNDS

TAKUJI ARAI

Department of Economics, Keio University

MASAAKI FUKASAWA

Department of Mathematics, Osaka University

We study convex risk measures describing the upper and lower bounds of a good deal bound, which is a subinterval of a no-arbitrage pricing bound. We call such a convex risk measure a good deal valuation and give a set of equivalent conditions for its existence in terms of market. A good deal valuation is characterized by several equivalent properties and in particular, we see that a convex risk measure is a good deal valuation only if it is given as a risk indifference price. An application to shortfall risk measure is given. In addition, we show that the no-free-lunch (NFL) condition is equivalent to the existence of a relevant convex risk measure, which is a good deal valuation. The relevance turns out to be a condition for a good deal valuation to be reasonable. Further, we investigate conditions under which any good deal valuation is relevant.

KEY WORDS: convex risk measure, good deal bound, Orlicz space, risk indifference price, fundamental theorem of asset pricing.

1. INTRODUCTION

The no-arbitrage framework in mathematical finance is not sufficient for providing a unique price for a given contingent claim in an incomplete market. Instead provided is only a no-arbitrage pricing bound. Because it is in general too wide to be useful in financial practice, needed is an alternative way to find nice candidates of prices of contingent claims. As a method to give a sharper pricing bound, the framework of no-good-deal has been discussed in much literature; for example, Arai (2011), Becherer (2009), Bernardo and Ledoit (2000), Björk and Slinko (2006), Carr et al. (2001), Černý and Hodges (2002), Černý (2003), Cochrane and Saá-Requejo (2000), Jaschke and Küchler (2001), Klöppel and Schweizer (2007), Larsen et al. (2005), and Staum (2004). The no-arbitrage pricing bound for a claim is obtained by excluding prices which enable either a seller or buyer to enjoy an arbitrage opportunity by trading the claim and selecting a suitable portfolio strategy. The price in a market should be consistent with this bound to make the market viable. On the other hand, an upper (resp. a lower) good deal bound may be interpreted as

The authors would like to thank Professors Freddy Delbaen and Martin Schweizer for their valuable comments and suggestions. This work was done when the authors were Visiting Professors of ETH Zurich. Takuji Arai was supported by Scientific Research (C) No.22540149 from the Ministry of Education, Culture, Sports, Science and Technology of Japan. Masaaki Fukasawa was supported by Japan Science and Technology Agency, CREST.

Manuscript received July 2011; final revision received August 2012.

Address correspondence to Takuji Arai, Department of Economics, Keio University, 2-15-45, Mita, Minato-ku, Tokyo, 1088345, Japan; e-mail: arai@econ.keio.ac.jp.

DOI: 10.1111/mafi.12020

determined by the seller's (resp. the buyer's) attitude to the risk associated with the claim. This can be considered as a generalization of both the pricing principle of no-arbitrage and exponential utility indifference valuation. Denote by $a(x)$ such an upper bound for a claim x. The functional a is supposed to have the following properties:

(i) $a(0) = 0$,
(ii) $a(x) \leq a(y)$ if $x \leq y$,
(iii) $a(x + c) = a(x) + c$ for any $c \in \mathbb{R}$,
(iv) $a(\lambda x + (1 - \lambda)y) \leq \lambda a(x) + (1 - \lambda)a(y)$ for any $\lambda \in [0, 1]$

for any claims x and y. In the second property, the inequality $x \leq y$ is in the almost sure sense, where we regard the claims as random variables. In the third, the element $c \in \mathbb{R}$ stands for a deterministic cash-flow. The last one represents the risk-aversion of the seller taking into account the impact of diversification. In brief, we suppose that ρ_a defined as $\rho_a(x) := a(-x)$ is a normalized convex risk measure. If we impose additionally the positive homogeneity: $a(\lambda x) = \lambda a(x)$ for all x and $\lambda \geq 0$, which implies the subadditivity: $a(x + y) \leq a(x) + a(y)$ for all x and y, then ρ_a becomes a coherent risk measure. By the same sort argument as above, a functional b which refers to a lower good deal bound is given by a normalized convex risk measure ρ_b as $b(x) = -\rho_b(x)$.

A good deal bound should be a subinterval of the no-arbitrage pricing bound, so not every convex risk measure yields a good deal bound. The aim of this paper is to characterize such a convex risk measure, which we call a good deal valuation (GDV hereafter); we define GDV as a normalized convex risk measure ρ with the Fatou property such that for any claim x, the value $\rho(-x)$ lies in the no-arbitrage pricing bound of x. This definition of GDV is given from sellers' viewpoint; for a GDV ρ and a claim x, $a(x) := \rho(-x)$ serves as an ask price of x. Nevertheless, it is easy to see that if ρ is a GDV, then $b := -\rho$ gives bid prices. We impose the Fatou property as a natural continuity condition for good deal bounds.

First we investigate equivalent conditions for the existence of a GDV. Among others, we show that a GDV exists under a condition weaker than the no-arbitrage one, which means that there may be GDVs even if the underlying market admits an arbitrage opportunity. Further we study equivalent conditions for a given ρ to be a GDV. In particular, we see that any GDV is given as a risk indifference price. The concept of risk indifference price has been undertaken by Xu (2006). There is much literature on this topic (Elliott and Siu 2010, Klöppel and Schweizer 2007, Øksendal and Sulem 2009, among others). Some of the above papers observe that a risk indifference price provides a good deal bound. Our assertion is that its reverse implication also holds true, which seems a new insight.

As mentioned before, GDV may exist even in markets with free lunch. We observe the equivalence between the no-free-lunch condition (NFL) and the existence of a relevant GDV, that is a relevant convex risk measure which is a GDV. This could be considered as a version of Fundamental Theorem of Asset Pricing (FTAP). Moreover as a version of Extension Theorem, we see that the relevance of a GDV is equivalent to that the extended market by the GDV satisfies NFL. We see also that the relevance is equivalent to the no-near-arbitrage condition (NNA) introduced by Staum (2004). We give an example (Example 4.9) which shows that NFL for the original market does not ensure NNA in general for a given GDV. We investigate conditions under which any GDV is relevant, and illustrate some examples related to this topic.

Now we mention the preceding results on FTAP from the viewpoint of good deal bound. Kreps (1981) introduced NFL and proved FTAP as well as Extension Theorem. Černý and Hodges (2002) established the framework of good deal bound and gave a version of Extension Theorem. Jaschke and Küchler (2001) showed that good deal bounds are essentially equivalent to coherent risk measures and gave a variant of FTAP. Staum (2004) extended their results to the noncoherent case. Bion-Nadal (2009) introduced a dynamic version and gave an associated FTAP. In Jaschke and Küchler (2001) and Staum (2004), an acceptance set reflecting an investor's preference is given first, and a convex risk measure induced by it is considered as a functional describing a good deal bound. Our approach is different, although we treat very similar problems. In our study, a convex risk measure is given first, and necessary and sufficient conditions for the given convex risk measure to be a GDV is discussed. This approach is in the same spirit as Bion-Nadal (2009). Our results provide a deeper understanding of a convex risk measure as a pricing functional in a market. Although our framework appears to be static, an extension to the dynamic framework of Bion-Nadal (2009) can be done in a straightforward manner. A detailed comparison with Staum (2004) and Bion-Nadal (2009) will be given in Remarks 3.5, 4.4, and 4.7.

In Section 2, we describe our model and prepare notation. In particular, we introduce the definitions and some basic properties of superhedging cost and risk indifference price. Main results are given in Sections 3 and 4.

2. PRELIMINARIES

Here, we introduce our framework and several basic results.

2.1. The Orlicz Space

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. The Orlicz space L^Ψ with Young function Ψ is defined as the set of the random variables X such that there exists $c > 0$,

$$\mathbb{E}[\Psi(cX)] < \infty.$$

Here, we call $\Psi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ a Young function if it is an even convex function with $\Psi(0) = 0$, $\Psi(x)\uparrow\infty$ as $x\uparrow\infty$ and $\Psi(x) < \infty$ for x in a neighborhood of 0. It is a Banach lattice with the gauge norm

$$\|X\| := \inf\{c > 0; \mathbb{E}[\Psi(X/c)] \leq 1\}$$

and pointwise ordering in the almost sure sense. In the case of $\Psi = \Psi_\infty$:

$$\Psi_\infty(x) := \begin{cases} 0 & \text{if } |x| \leq 1, \\ \infty & \text{otherwise} \end{cases}$$

we have $L^\Psi = L^\infty$. Further, for $\Psi_p(x) := |x|^p$ with $p \geq 1$, we have $L^{\Psi_p} = L^p$. The Orlicz heart M^Ψ is a subspace of L^Ψ defined as

$$M^\Psi := \{X \in L^\Psi | \mathbb{E}[\Psi(cX)] < \infty \text{ for all } c > 0\}.$$

In this paper, we consider the set of the future cash-flows L to be either L^Ψ or M^Ψ with a fixed Young function Ψ. This specification would be justified by noting that L becomes a linear space of random variables with natural ordering and sufficiently abstract in that it incorporates L^p spaces with $1 \leq p \leq \infty$. More importantly, a Young function Ψ may be connected to a utility function u as $\Psi(x) = -u(-|x|))$ and then L becomes a suitable space where expected utility maximization is considered (see, e.g., Biagini and Frittelli 2009). Note that the case of exponential utility is covered. Our treatment and results do not depend on a specific choice of Ψ. This generality is indeed necessary to derive a conclusion which does not depend on a specific choice of utility function.

Let $M \subset L$ be the set of the 0-attainable claims. Each element of M represents a future payoff which investors can super-replicate with 0 initial endowment. Simultaneously, M might be regarded as the set of strategies which investors can take. We suppose that M is a convex cone including L_-, where we denote L_+ (resp. L_-):$=$ $\{x \in L | x \geq 0$ (resp. $\leq$)$\}$. Although M is defined as a subset of L^∞ in much literature, we use the Orlicz space framework, because it enables us to treat more various models, say, the exponential hedging introduced in Corollary 3.12 of Biagini et al. (2011). Besides, assumed on M is the minimal property that the set of the 0-attainable claims possesses. This enables us to treat such models taking transaction costs into account as Example 4.11 at the end of the paper.

EXAMPLE 2.1. Here we see examples of M appeared in the preceding studies of frictionless markets. In such an idealized market, M is described by the sum of L_- and a set of stochastic integrals with respect to a semimartingale representing underlying asset price processes. This set of stochastic integrals is a convex cone if so is the set of the admissible strategies. Let S be the underlying asset price process being an $\mathbf{R}^d$-valued semimartingale defined on $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}_{t\in[0,T]})$, where $\{\mathcal{F}_t\}_{t\in[0,T]}$ is a filtration with the usual conditions. Then M is typically given in the form

$$M = \left(\left\{ \int_0^T H_t \mathrm{d}S_t \,\middle|\, H \in \mathcal{H} \right\} - L^0_+ \right) \cap L, \tag{2.1}$$

where $\mathcal{H}$ is the set of the admissible strategies and L^0_+ is the set of the nonnegative random variables.

1. (Section 5 of Delbaen and Schachermayer 2006) Let $\mathcal{H}^1$ be the set of processes H of the form $H_t = \sum_{i=1}^n h_i 1_{(\tau_{i-1},\tau_i]}(t)$, where $0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_n \leq T$ are stopping times and for each i, h_i is an $\mathcal{F}_{\tau_{i-1}}$-measurable random variable such that the stopped process S^{τ_n} and $h_1, \ldots, h_n$ are bounded. Because $\mathcal{H}^1$ is a convex cone, so is M defined by (2.1) with $\mathcal{H} = \mathcal{H}^1$.
2. Letting $\mathcal{H}^2$ be the set of S-integrable predictable processes such that $\int_0^t H_s \mathrm{d}S_s$ is uniformly bounded from below, M defined by (2.1) with $\mathcal{H} = \mathcal{H}^2$ is a convex cone because so is $\mathcal{H}^2$. It seems that $\mathcal{H}^2$ is reasonable as the set of admissible strategies as explained in Section 8 of Delbaen and Schachermayer (2006).
3. Note that $\mathcal{H}^2$ may be reduced to $\{0\}$ when S is not necessarily locally bounded. As a natural framework for such cases, we can consider W-admissible strategies as in Biagini et al. (2011). Fix $W \in L$ with $W \geq 1$, and denote by $\mathcal{H}^3$ the set of S-integrable predictable processes H such that there exists a constant $c > 0$ satisfying $\int_0^t H_s \mathrm{d}S_s \geq -cW$ for any $t \in [0, T]$. Then, $\mathcal{H}^3$ is a convex cone and so, M defined by (2.1) with $\mathcal{H} = \mathcal{H}^3$ also forms a convex cone.

Let L^*_+ be the set of all positive linear functionals on L. Remark that any element of L^*_+ is continuous by the Namioka-Klee theorem (see Biagini and Frittelli 2009 for an extended result). Both the cases of $L = L^\Psi$ and $L = M^\Psi$ are treated in a unified way in the following. Let $L^\dagger := L^{\Psi^\dagger}$, where $\Psi^\dagger$ is the complementary function of Ψ defined as

$$\Psi^\dagger(y) := \sup_{x\in\mathbb{R}}\{xy - \Psi(x)\}.$$

Define a set of probability measures $\mathcal{P} := \{Q \ll \mathbb{P} | dQ/d\mathbb{P} \in L^\dagger\}$. Further, let $\overline{L}^* := \{g \in L^*_+ | g(1) = 1, g(m) \le 0 \text{ for any } m \in M\}$, $\mathcal{Q} := \{Q \in \mathcal{P} | dQ/d\mathbb{P} \in \overline{L}^*\}$, and $\mathcal{Q}^e := \{Q \in \mathcal{Q} | Q \sim \mathbb{P}\}$. For $Q \in \mathcal{P}$, denote by $\mathbb{E}_Q$ the corresponding expectation operator. By Young's inequality:

$$\frac{xy}{ab} \le \Psi\left(\frac{x}{a}\right) + \Psi^\dagger\left(\frac{y}{b}\right)$$

for any $x, y \in \mathbb{R}$ and $a, b > 0$, the operation $\mathbb{E}_Q$ enables us to identify $\mathcal{P}$ with a subset of L^*_+.

2.2. Convex Risk Measure

Here, we collect several notions and results on convex risk measures which we utilize in this paper. A convex risk measure ρ is a $(-\infty, +\infty]$-valued functional on L satisfying

properness: $\rho(0) < \infty$,
monotonicity: $\rho(x) \ge \rho(y)$ if $x \le y$,
cash-invariance: $\rho(x + c) = \rho(x) - c$ for any $c \in \mathbb{R}$,
convexity: $\rho(\lambda x + (1 - \lambda)y) \le \lambda\rho(x) + (1 - \lambda)\rho(x)$ for any $\lambda \in [0, 1]$,

for any $x, y \in L$. A convex risk measure ρ is a **coherent risk measure** if it satisfies in addition,

positive homogeneity: $\rho(cx) = c\rho(x)$ for any $x \in L$ and any $c > 0$.

THEOREM 2.2. (Biagini and Frittelli 2009) *Let ρ be a convex risk measure. Then,*

$$\rho(-x) = \max_{g\in L^*_+, g(1)=1}\{g(x) - \rho^*(g)\}$$

for $x \in \mathrm{Int}\{\rho < \infty\}$, where for $g \in L^*_+$,

$$\rho^*(g) := \sup_{x\in L}\{g(x) - \rho(-x)\}.$$

A convex risk measure ρ is said to have **the Fatou property** if for any increasing sequence $\{x_n\}\subset L$ with $x_n \uparrow x_\infty$ a.s., $\rho(-x_n)\uparrow\rho(-x_\infty)$. Denote by $\mathcal{R}$ the set of all convex risk measures with $\rho(0) = 0$ and the Fatou property.

THEOREM 2.3. (Biagini and Frittelli 2009) *For $\rho \in \mathcal{R}$, we have for $x \in L$,*

$$\rho(x) = \sup_{Q\in\mathcal{P}}\{\mathbb{E}_Q[-x] - \rho^*(Q)\}. \tag{2.2}$$

A convex risk measure ρ is said to be **finite** *if $\rho(x) < \infty$ for all $x \in L$.*

REMARK 2.4. In the case of $L = M^\Psi$, it is known that $L^\dagger$ coincides with the dual of L and the supremum in (2.2) is attained. Moreover, every finite convex risk measure has the

Fatou property. See Biagini and Frittelli (2009) for the detail. The finiteness condition cannot be dropped as we see in Example 2.9 below. If Ψ satisfies the Δ_2 condition: there exist $t_0 > 0$ and $K > 0$ such that $\Psi(2t) \leq K\Psi(t)$ for any $t \geq t_0$, then we have $L^\Psi = M^\Psi$. For $p \in [1, \infty)$, L^p is an example of such cases.

A convex risk measure ρ is said to have **the Lebesgue property** if for any sequence $\{x_n\} \subset L$ with $\sup_n \|x_n\|_\infty < \infty$ and $x_n \to x_\infty$ a.s., it holds that $\rho(x_n) \to \rho(x_\infty)$ as $n \to \infty$. Here $\|\cdot\|_\infty$ refers to the L^∞ norm. This definition was introduced in Jouini et al. (2006) for the $L = L^\infty$ case. Because any continuous linear functional on L can be decomposed into the sum of an element of $L^\dagger \subset L^1$ and a purely finitely additive signed measure (see Rao and Ren 1991), the same argument as the proof of Theorem 2.4 in Jouini et al. (2006) can apply to have the following result with the aid of Theorem 2.2 above.

THEOREM 2.5. *For a finite convex risk measure ρ, the following are equivalent:*

1. *ρ has the Lebesgue property.*
2. *for any $\alpha > 0$ and a sequence of measurable sets A_n with $P(A_n) \to 0$, it holds that $\rho(-\alpha 1_{A_n}) \to 0$ as $n \to \infty$.*
3. *for any $c > 0$, the set $\{g \in L_+^*; \rho^*(g) \leq c\}$ is a uniformly integrable subset of $L^\dagger$ and for any $x \in L$, it holds that*

$$\rho(-x) = \max_{Q \in \mathcal{P}} \{\mathbb{E}_Q[x] - \rho^*(Q)\}. \tag{2.3}$$

Note that the Fatou property follows from the Lebesgue property by (2.3).

A convex risk measure is said to be **relevant** if $\rho(-z) > 0$ for any $z \in L_+ \backslash \{0\}$. The relevance was introduced in Delbaen (2002) as a condition for coherent risk measures with the Fatou property to be represented as (2.2) with a set of equivalent probability measures instead of $\mathcal{P}$.

2.3. Superhedging Cost

Here we discuss superhedging cost. Define a functional ρ^0 on L as

$$\rho^0(x) := \inf\{c \in \mathbb{R} \mid \text{there exists } m \in M \text{ such that } c + m + x \geq 0\}. \tag{2.4}$$

Because $\rho^0(-x)$ represents the superhedging cost for a claim x, it gives the upper no-arbitrage pricing bound for x. In fact if a seller could sell x with a price greater than $\rho^0(-x)$, then she could enjoy an arbitrage opportunity by taking a suitable strategy from M. By the same reasoning the lower no-arbitrage pricing bound for x is given by $-\rho^0(x)$.

LEMMA 2.6. *The superhedging cost ρ^0* is $(-\infty, \infty]$-*valued if and only if $\overline{L}^* \neq \emptyset$. If ρ^0* is $(-\infty, \infty]$-*valued, then it is a coherent risk measure with*

$$(\rho^0)^*(g) = \begin{cases} 0 & if\ \ g \in \overline{L}^*, \\ \infty & otherwise. \end{cases}$$

Proof. Suppose that $\overline{L}^* \neq \emptyset$. If there exists $x \in L$ with $\rho^0(x) = -\infty$, then (2.4) implies that for any $c > 0$, we can find $m^c \in M$ such that $-c + m^c + x \geq 0$. This gives $g(x) \geq c$, so that $g(x) = \infty$ for any $g \in \overline{L}^*$. This is a contradiction, so ρ^0 is $(-\infty, \infty]$-valued. Next, suppose that $\overline{L}^* = \emptyset$. Then there exists a sequence $\{m_n\} \subset M$ such that $\|m_n - 1\| \to 0$ as $n \to \infty$. In fact if the closure M^s of M does not include 1, then the Hahn-Banach theorem implies the existence of a continuous linear functional μ such that $\mu(1) > \sup_{m \in M^s} \mu(m)$. The RHS is 0 because M^s is a cone. That $L_- \subset M^s$ implies $\mu \in L^*_+$. This means $\overline{L}^* \neq \emptyset$, which is a contradiction. Now, taking a subsequence if necessary, we may suppose that $\sum_{n=1}^\infty \|m_n - 1\| < \infty$. Then for $x := \sum_{n=1}^\infty |m_n - 1| \in L$ and for all $N \in \mathbb{N}$,

$$x \geq \sum_{n=1}^{N} (1 - m_n) = N - \sum_{n=1}^{N} m_n,$$

which implies that $\rho^0(x) \leq -N$, and so $\rho^0(x) = -\infty$.

Now we see that ρ^0 is a coherent risk measure and calculate $(\rho^0)^*$. The convexity and positive homogeneity of ρ^0 follow from the assumption that M is a convex cone. The monotonicity and cash-invariance are obvious. The fact that $\rho^0(0) \leq 0$ implies that $(\rho^0)^*(g) \geq 0$ for any $g \in L^*_+$. On the other hand, for any $\varepsilon > 0$ and $x \in L$, we can find $m^\varepsilon \in M$ so that $\rho^0(x) + \varepsilon + m^\varepsilon + x \geq 0$. Because $g(m^\varepsilon) \leq 0$ for $g \in \overline{L}^*$, we have $\rho^0(x) + \varepsilon \geq g(-x)$, which implies that

$$\sup_{x \in L} \{g(-x) - \rho^0(x)\} \leq 0.$$

We therefore have $(\rho^0)^*(g) = 0$ for $g \in \overline{L}^*$. For $g \in L^*_+ \setminus \overline{L}^*$, there exists $m \in M$ such that $g(m) > 0$. Because M is a cone,

$$(\rho^0)^*(g) = \sup_{x \in L} \{g(-x) - \rho^0(x)\} \geq \sup_{m \in M} \{g(m) - \rho^0(-m)\} \geq \sup_{m \in M} g(m) = \infty.$$

□

For later use, we define for $x \in L$,

$$\widehat{\rho^0}(x) := \begin{cases} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[-x] & \text{if } \mathcal{Q} \neq \emptyset \\ -\infty & \text{otherwise.} \end{cases}$$

By definition $\widehat{\rho^0}$ is a coherent risk measure on L belonging to $\mathcal{R}$ if $\mathcal{Q} \neq \emptyset$.

LEMMA 2.7. *If $\mathcal{Q} \neq \emptyset$, then $-\rho^0(x) \leq -\widehat{\rho^0}(x) \leq \widehat{\rho^0}(-x) \leq \rho^0(-x)$ for any $x \in L$. Moreover if $\mathcal{Q}^e \neq \emptyset$, then $\widehat{\rho^0}$ is relevant.*

Proof. For any $x \in L$ and $\varepsilon > 0$, there exists $m \in M$ such that $\rho^0(x) + \varepsilon + m + x \geq 0$. Then we have $\mathbb{E}_Q[-x] \leq \rho^0(x) + \varepsilon$ for any $Q \in \mathcal{Q}$. Because $Q \in \mathcal{Q}$ and $\varepsilon > 0$ are arbitrary, we have $\widehat{\rho^0}(x) \leq \rho^0(x)$. It suffices then to observe that $\widehat{\rho^0}(x) + \widehat{\rho^0}(-x) \geq 2\widehat{\rho^0}(0) = 0$ by the convexity.

The relevance under $\mathcal{Q}^e \neq \emptyset$ is shown by noting that

$$\widehat{\rho^0}(-x) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[x] = \sup_{Q \in \mathcal{Q}^e} \mathbb{E}_Q[x].$$

In fact if there exists $Q_1 \in \mathcal{Q}$ with $\mathbb{E}_{Q_1}[x] > \sup_{Q \in \mathcal{Q}^e} \mathbb{E}_Q[x]$, then we have a contradiction because for any $Q_0 \in \mathcal{Q}^e$, $\lambda Q_0 + (1 - \lambda) Q_1 \in \mathcal{Q}^e$ converges to Q_1 in $\sigma(L^\dagger, L)$ as $\lambda \downarrow 0$. □

REMARK 2.8.

1. As theorem 9.5.8 of Delbaen and Schachermayer (2006) stated, we have $\rho^0 = \widehat{\rho^0}$ if M is defined as a subset of L^∞ induced by a bounded semimartingale satisfying the NFLVR.
2. Theorem 5 of Biagini and Frittelli (2004) mentioned a similar result to Lemma 2.7 although their setting is different from ours.
3. Proposition 4.2 of Biagini et al. (2011) introduced results on utility indifference price in the Orlicz space framework and, in addition, mentioned relationship to superhedging cost.

The following example shows that ρ^0 does not necessarily coincides with $\widehat{\rho^0}$, so is not always represented as (2.2) even though $\mathcal{Q}$ is not empty.

EXAMPLE 2.9. Let $L = L^p$ with $p \in [1, \infty)$ and take the following set as M:

$$M = \{-z + \mathbb{E}_{Q^0}[z] \mid z \in L^\infty\} - L_+,$$

where $Q^0 \in \mathcal{P}$ is arbitrarily fixed. Any element of M is bounded from above. Therefore by the definition of ρ^0, we have $\rho^0(-z) = \infty$ for $z \in L_+$ which is not bounded from above. It is clear that $Q^0 \in \mathcal{Q}$, so that $\overline{L}^* \neq \emptyset$. Therefore ρ^0 is a coherent risk measure by Lemma 2.6. Moreover $\mathcal{Q} = \{Q^0\}$ because for any $Q \in \mathcal{Q}$, we have $\mathbb{E}_{Q^0}[z] \leq \mathbb{E}_Q[z]$ for any $z \in L^\infty$, which implies that $Q = Q^0$. Therefore ρ^0 cannot be represented as (2.2).

In fact we can prove that ρ^0 does not have the Fatou property. Let $z \in L_+$ be unbounded from above. Consider the increasing sequence $z_n = z \wedge n$, $n \in \mathbb{N}$. Because $n - z_n \in L^\infty$, we have $z_n - \mathbb{E}_{Q^0}[z_n] \in M$. It follows that $\rho^0(-z_n) \leq \mathbb{E}_{Q^0}[z_n] \to E_{Q^0}[z] < \infty$, whereas $\rho^0(-z) = \infty$.

Remark that M of this example can be represented in terms of the familiar Black-Scholes model. Suppose an asset price process S follows $\mathrm{d}S_t = \sigma S_t \mathrm{d}W_t$ under Q^0, where $\sigma > 0$ and W is a Brownian motion which generates the σ-field $\mathcal{F}$. Let $\mathcal{H}$ be the set of S-integrable processes H such that $\int_0^\cdot H_t \mathrm{d}S_t$ are bounded. Then, by the Itô representation theorem, M coincides with (2.1).

2.4. Risk Indifference Prices

Here we recall risk indifference price. Given a convex risk measure ρ, define a functional $I(\rho)$ on L as

$$\begin{aligned} I(\rho)(x) &:= \inf\{c \in \mathbb{R} \mid \inf_{m\in M} \rho(c+m+x) \leq \inf_{m\in M} \rho(m)\} \\ &= \inf\{c \in \mathbb{R} \mid \inf_{m\in M} \rho(m+x) - c \leq \inf_{m\in M} \rho(m)\}. \end{aligned} \tag{2.5}$$

Then $I(\rho)(-x)$ describes the risk indifference seller's price for x induced by ρ as introduced in Xu (2006). The idea is explained as follows. If a trader sells a claim x with a price $c > I(\rho)(-x)$, then she can find $\widehat{m} \in M$ such that $\rho(c + \widehat{m} - x) \leq \inf_{m\in M} \rho(m)$. This means that selling the claim with the price does not increase the risk measured by ρ. The following lemma gives a representation of $I(\rho)$. Denote $\check{\rho} := \rho - \inf_{m\in M} \rho(m)$.

LEMMA 2.10. *Let ρ be a convex risk measure. If $I(\rho)$ is $(-\infty, \infty]$-valued, then we have* $\inf_{m\in M} \rho(m) \in \mathbb{R}$ *and that $I(\rho)$ is a convex risk measure with*

$$I(\rho)^*(g) = \begin{cases} \check{\rho}^*(g) = \rho^*(g) + \inf_{m\in M} \rho(m), & if\ g \in \overline{L}^* \\ \infty & otherwise. \end{cases}$$

If $I(\rho) \in \mathcal{R}$ in addition, then $\mathcal{Q} \neq \emptyset$ and

$$I(\rho)(x) = \sup_{Q\in\mathcal{Q}} \{\mathbb{E}_Q[-x] - \check{\rho}^*(Q)\}. \tag{2.6}$$

Proof. Because $\rho(0) < \infty$ and $0 \in M$, we have $I(\rho)(0) = 0$ or $-\infty$ depending on whether $\inf_{m\in M} \rho(m)$ is finite or $-\infty$. Therefore if $I(\rho) > -\infty$ then $\inf_{m\in M} \rho(m)$ is finite and $I(\rho)(x) = \inf_{m\in M} \rho(x + m) - \inf_{m\in M} \rho(m) = \inf_{m\in M} \check{\rho}(x + m)$. From this the cash-invariance and monotonicity of $I(\rho)$ are obvious. The convexity follows from that M is convex. Because M is a cone, we have

$$\begin{aligned} I(\rho)^*(g) &= \sup_{x\in L} \{g(-x) - I(\rho)(x)\} \\ &= \sup_{m\in M} \sup_{x\in L} \{g(-x) - \check{\rho}(x + m))\} \\ &= \sup_{m\in M} \{g(m) + \check{\rho}^*(g)\} \\ &= \begin{cases} \check{\rho}^*(g) & \text{if } g \in \overline{L}^* \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

By Theorem 2.3, we have (2.6) if $I(\rho) \in \mathcal{R}$ and in particular, $\mathcal{Q} \neq \emptyset$. □

3. GOOD DEAL VALUATIONS

In this section, we discuss conditions under which a convex risk measure yields a good deal bound. A good deal bound should be a subinterval of the no-arbitrage pricing bound. We therefore introduce the following definition.

DEFINITION 3.1. *A convex risk measure $\rho \in \mathcal{R}$ is said to be a GDV if*

$$\rho(-x) \in [-\rho^0(x), \rho^0(-x)] \text{ for any } x \in L. \tag{3.1}$$

As mentioned in the Introduction, the above definition is given from seller's viewpoint. Nevertheless, (3.1) is equivalent to

$$-\rho(x) \in [-\rho^0(x), \rho^0(-x)] \text{ for any } x \in L,$$

which is from buyer's viewpoint. In addition, $-\rho(x) \leq \rho(-x)$ for any $x \in L$ because $\rho(x) + \rho(-x) \geq 2\rho(0) = 0$ by the convexity. For a GDV ρ, a good deal bound may be constructed as $[-\rho(x), \rho(-x)]$, which is a subinterval of $[-\rho^0(x), \rho^0(-x)]$. Note that the upper and lower bounds of a good deal bound may be described by different GDVs.

3.1. Existence of Good Deal Valuations

Here, we present a set of equivalent conditions for the existence of a GDV. Denote by $\overline{M}$ the closure of M in $\sigma(L, L^\dagger)$.

THEOREM 3.2. *The following are equivalent:*

1. $\mathcal{Q} \neq \emptyset$.
2. *There exists a GDV.*
3. $\mathbb{P}(m > 0) < 1$ *for any* $m \in \overline{M}$.
4. $1 \notin \overline{M}$.

Proof. 1⇒2: This is from Lemma 2.7.

2⇒1: Let ρ be a GDV. Because $\rho(-m) \le \rho^0(-m) \le 0$ for any $m \in M$,

$$\rho^*(Q) = \sup_{x \in L}\{\mathbb{E}_Q[-x] - \rho(x)\} \ge \sup_{m \in M}\{\mathbb{E}_Q[m] - \rho(-m)\} \ge \sup_{m \in M} \mathbb{E}_Q[m].$$

Then the cone property of M implies that $\rho^*(Q) = +\infty$ for any $Q \in \mathcal{P}\backslash\mathcal{Q}$. If $\mathcal{Q}$ is empty, then ρ equals to $-\infty$ identically by (2.2), which contradicts $\rho \in \mathcal{R}$.

1⇒3: If there exists $m \in \overline{M}$ such that $P(m > 0) = 1$, then we have $\mathbb{E}_Q[m] > 0$ for any $Q \in \mathcal{P}$, and so $\mathcal{Q} = \emptyset$.

3⇒4: This holds true clearly.

4⇒1: Because $1 \notin \overline{M}$, the Hahn-Banach theorem implies that there exists $z \in L^\dagger$ such that

$$\sup_{m \in \overline{M}} \mathbb{E}[zm] < \mathbb{E}[z]. \tag{3.2}$$

We have $\sup_{m \in \overline{M}} \mathbb{E}[zm] = 0$ because $0 \in M$ and $\overline{M}$ is a cone. Because $L_- \subset \overline{M}$, we have then that $z \in L^*_+ \cap L^\dagger$, so that $z/\mathbb{E}[z] \in \mathcal{Q}$. □

Condition 3 in the above theorem is weaker than the no-arbitrage condition. This means that a GDV may exist even if there is an arbitrage opportunity. The following example shows that we cannot replace $\overline{M}$ with M in Conditions 3 and 4.

EXAMPLE 3.3. We take the Lebesgue measure space on (0, 1] as the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let u be the random variable given by $u(\omega) := \omega$, and M be given by $\{cu | c \ge 0\} - L_+$. We can see several interesting facts on this example as follows:

1. We consider the following two conditions:
   (a) $\mathbb{P}(m > 0) < 1$ for any $m \in M$,
   (b) $1 \notin M$.
   This example satisfies (b), but does not satisfy (a). Replacing M by $\overline{M}$, the two conditions become equivalent by Theorem 3.2.
2. Because $1 \notin M$, we have $\rho^0(0) = 0$. Therefore if we take $L = L^\infty$, then ρ^0 is a finite coherent risk measure. In fact for any $x \in L^\infty$, $-\|x\|_\infty = \rho^0(\|x\|_\infty) \le \rho^0(x) \le \rho^0(-\|x\|_\infty) = \|x\|_\infty$ by monotonicity. On the other hand, ρ^0 is not a convex risk measure on $L = L^p$ with $p \in [1, \infty)$ since $\overline{L}^* = \mathcal{Q}$ is empty. Note that for $x(\omega) := \log \omega$, we have $\rho^0(-x) = -\infty$.
3. Notice that $\mathcal{Q}$ is empty despite that the above Condition (b) holds. We therefore need to take the closure of M in Condition 4 of Theorem 3.2. In fact, considering

the sequence $m_n := (nu) \wedge 1$, m_n converges to 1, and so this example does not satisfy Conditions 3 nor 4.

3.2. Equivalent Conditions for Good Deal Valuations

Here we present conditions for a given ρ to be a GDV. The main contribution of the following theorem is to show the equivalence between GDVs and risk indifference prices.

THEOREM 3.4. *For any $\rho \in \mathcal{R}$, the following conditions are equivalent:*

1. *ρ is a GDV.*
2. *$\rho(-m) \leq 0$ for any $m \in M$.*
3. *There exists a function $c : \mathcal{Q} \to \mathbb{R}$ such that for any $x \in L$,*

$$\rho(x) = \sup_{Q\in\mathcal{Q}} \{\mathbb{E}_Q[-x] - c(Q)\}.$$

4. *There exists $\eta \in \mathcal{R}$ such that $\rho = I(\eta)$.*

4′. *$\rho = I(\rho)$, that is, ρ is a fixed point of I.*

5. *$\rho(-x) \in [-\widehat{\rho^0}(x), \widehat{\rho^0}(-x)]$ for any $x \in L$.*
6. *$\{\rho^0 \leq 0\} \subset \{\rho \leq 0\}$.*
7. *$\mathcal{Q} \supset \{Q \in \mathcal{P} | \rho^*(Q) < +\infty\}$.*
8. *There exists a convex set $A \subset L$ including 0 with $A + L_+ \subset A$ such that for any $x \in L$,*

$$\rho(x) = \inf\{c \in \mathbb{R} | \text{ there exists } m \in M \text{ such that } c + m + x \in A\}. \tag{3.3}$$

Proof. 1⇒2: This is because $\rho(-m) \leq \rho^0(-m) \leq 0$ for any $m \in M$ by the definitions of GDV and ρ^0.

2⇒7: We have

$$\rho^*(Q) = \sup_{x\in L}\{\mathbb{E}_Q[-x] - \rho(x)\} \geq \sup_{m\in M}\{\mathbb{E}_Q[m] - \rho(-m)\} \geq \sup_{m\in M} \mathbb{E}_Q[m].$$

Because M is a cone, we have $\rho^*(Q) = \infty$ for any $Q \in \mathcal{P} \setminus \mathcal{Q}$.

7⇒3: This is from Theorem 2.3.

3⇒4′ and 4: Because $\rho \in \mathcal{R}$, we have

$$\rho(-m) = \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[m] - c(Q)\} \leq -\inf_{Q\in\mathcal{Q}} c(Q) = \rho(0) = 0$$

for any $m \in M$. Then, by the convexity, we have $\rho(m) + \rho(-m) \geq 2\rho(0) = 0$ and so, $\inf_{m\in M} \rho(m) = 0$. Therefore,

$$I(\rho)(x) = \inf_{m\in M} \rho(m + x) - \inf_{m\in M} \rho(m) \leq \rho(x) \tag{3.4}$$

and

$$I(\rho)(x) = \inf_{m\in M} \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[-m - x] - c(Q)\} \geq \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[-x] - c(Q)\} = \rho(x).$$

4⇒5: By Lemma 2.10, $\rho = I(\eta)$ is represented as

$$\rho(x) = \sup_{Q\in\mathcal{Q}} \left\{ \mathbb{E}_Q[-x] - \check{\eta}^*(Q) \right\}.$$

Because $\rho(0) = 0$, we have $\check{\eta}^*(Q) \geq 0$. Therefore,

$$\widehat{\rho^0}(-x) = \sup_{Q\in\mathcal{Q}} \mathbb{E}_Q[x] \geq \sup_{Q\in\mathcal{Q}} \left\{ \mathbb{E}_Q[x] - \check{\eta}^*(Q) \right\} = \rho(-x)$$

for all $x \in L$. It suffices then to recall that $\rho(x) + \rho(-x) \geq 2\rho(0) = 0$ by the convexity.

5⇒1: This is from Lemma 2.7.

3⇒6: For any $x \in \{\rho^0 \leq 0\}$, Lemma 2.7 implies that $\sup_{Q\in\mathcal{Q}} \mathbb{E}_Q[-x] = \widehat{\rho^0}(x) \leq 0$. We have then

$$\rho(x) = \sup_{Q\in\mathcal{Q}} \{\mathbb{E}_Q[-x] - c(Q)\} \leq \sup_{Q\in\mathcal{Q}} \{-c(Q)\} = \rho(0) = 0.$$

6⇒2: This is because $\rho^0(-m) \leq 0$ by definition.

4′⇒8: Taking $A = \{\rho \leq 0\}$ and noting that $\inf_{m\in M} \rho(m) = 0$, we have

$$\begin{aligned}
\rho(x) &= I(\rho)(x) = \inf_{m\in M} \rho(m+x) = \inf \left\{ c \in \mathbf{R} \middle| \inf_{m\in M} \rho(m+x) \leq c \right\} \\
&\leq \inf\{c \in \mathbf{R} |\ \text{there exists } m \in M \text{ such that } \rho(m+x) \leq c\} \\
&= \inf\{c \in \mathbf{R} |\ \text{there exists } m \in M \text{ such that } c + m + x \in A\} \\
&\leq \inf\{c \in \mathbf{R} | c + x \in A\} = \rho(x).
\end{aligned}$$

8⇒2: This is obvious. □

REMARK 3.5. Staum (2004) studied a very similar problem to ours. Here, we introduce his results roughly and compare his theorem 6.1 with theorem 3.4. Let A be a given acceptance set, that is, a nonempty subset of L such that $A + L_+ \subset A$. For simplicity, we assume additionally that A is convex and includes 0, although Staum (2004) did not impose it. Denoting the RHS of (3.3) by ρ_A, and assuming the Fatou property of ρ_A, it is a convex risk measure represented as $\rho_A(x) = \sup_{Q\in\mathcal{P}}\{E_Q[-x] - \rho_A^*(Q)\}$. The representation of GDV as ρ_A is important in that it implies robustness of GDV to quantitative specification of investor's risk preference. Now, we see the equivalence among (3.1), $\rho_A(0) = 0$ and the no-cashout condition (NC) introduced in Staum (2004): $\rho_A(-x) \geq -\rho^0(x)$ for any $x \in L$. In fact for any $x \in L$,

$$\begin{aligned}
\rho^0(x) &= \inf\{c \in \mathbb{R} |\ \text{there exists } m \in M \text{ such that } c + m + x \in L_+\} \\
&\geq \inf\{c \in \mathbb{R} |\ \text{there exists } m \in M \text{ such that } c + m + x \in A\} \\
&= \rho_A(x),
\end{aligned}$$

that is, the upper estimate for $\rho_A(-x)$ holds automatically. Thus, NC is equivalent to (3.1). The convexity of ρ_A implies that NC is equivalent to $\rho_A(0) = 0$. As a result, we can rewrite and extend theorem 6.1 of Staum (2004) as follows: The following are equivalent:

1. (Consistent pricing kernel) There exists a $g \in \overline{L}^*$ such that $\rho_A^*(g) = 0$.
2. ρ_A satisfies NC.
3. ρ_A satisfies (3.1).
4. $\rho_A(0) = 0$.

In other words, 1 and 2 are equivalent to

5. ρ_A is a GDV.

Note that theorem 6.1 showed only the equivalence between 1 and 2. Staum (2004) treated only ρ_A, which is a special class of convex risk measures and is not necessarily normalized, as a functional describing a good deal bound. On the other hand, Theorem 3.4 discusses equivalent conditions to be a GDV for general normalized convex risk measures. Furthermore, it also shows that ρ_A is the only class giving good deal bounds.

In this paper, we discuss GDVs by postulating the normalization. This provides us many advantages. We notice through the above discussion that, if $\rho_A(0) = 0$, then ρ_A is a GDV, that is, $\rho_A(0) = 0$ is a sufficient condition which we can check easily. Moreover, even if ρ_A is not a GDV, we can construct a new GDV by shifting ρ_A as $\rho_A - \rho_A(0)$. In summary, our discussion also gives a way to construct a GDV. On the other hand, it is not easy to check if a consistent pricing kernel exists. In addition, Staum (2004) did not show how to construct an alternative GDV when ρ_A is not a GDV.

REMARK 3.6. We consider the case where $\Psi(x) = e^{|x|} - 1$. Let M^0 be the set of replicatable claims with 0 initial cost. Thus, M is described by $M = M^0 + L^\Psi_-$. Suppose that M^0 is a subset of L^Ψ. Enlarging $L = L^\Psi$ to $L^p (p \geq 1)$, M is also extended to $M^0 + L^p_-$. However, $\rho^0(-x)$ is not influenced for any $x \in L^\Psi$, because we have $\rho^0(-x) = \inf\{c \in \mathbb{R} |$ there exists $m \in M^0$ such that $c + m - x \geq 0\}$. On the other hand, this enlargement may make the value of $\widehat{\rho^0}(-x)$ smaller. Hence, we can catch more GDVs by using L^Ψ as the underlying space rather than L^p.

As mentioned in the Introduction, many papers (Elliott and Siu 2010, Klöppel and Schweizer 2007, Øksendal and Sulem 2009, Xu 2006) treated risk indifference prices and some of them showed that a risk indifference price yields a good deal bound. On the other hand, Theorem 3.4 showed that a GDV is always a risk indifference price. It, therefore, supports the use of the operator I in constructing a good deal bound. We utilized, however, that a GDV has the Fatou property by definition. It should be noted that $I(\rho)$ does not necessarily have the Fatou property even if $\rho \in \mathcal{R}$. In other words, the operation does not necessarily preserve the Fatou property (see Example 3.9 below). Now we remark that it preserves the Lebesgue property that also could be regarded as a natural continuity requirement for good deal bounds as well as the Fatou property.

PROPOSITION 3.7. *Let ρ be a finite convex risk measure with the Lebesgue property and suppose that there exists $Q^0 \in \mathcal{Q}$ such that $\rho^*(Q^0) < \infty$. Then, $I(\rho)$ is a finite GDV with the Lebesgue property.*

Proof. By Theorem 2.5 and the existence of $Q^0 \in \mathcal{Q}$ such that $\rho^*(Q^0) < \infty$, we have, for any $x \in L$ and $m \in M$,

$$\begin{aligned}\rho(x+m) &= \max_{Q\in\mathcal{P}}\{E_Q[-x-m] - \rho^*(Q)\} \geq E_{Q^0}[-x-m] - \rho^*(Q^0)\\ &\geq E_{Q^0}[-x] - \rho^*(Q^0) > -\infty.\end{aligned}$$

Therefore $I(\rho)$ is $(-\infty, \infty]$-valued by (2.5), and so it is a convex risk measure by Lemma 2.10. Because ρ is finite, so is $I(\rho)$ by (2.5). Moreover for any $m \in M$, we have

$$I(\rho)(-m) = \inf_{m'\in M}\rho(-m+m') - \inf_{m'\in M}\rho(m') \leq \inf_{m'\in M}\rho(m') - \inf_{m'\in M}\rho(m') = 0. \tag{3.5}$$

Therefore by Theorem 3.4, it only remains to show that $I(\rho)$ has the Fatou property. By (2.3), it suffices to see that $I(\rho)$ has the Lebesgue property. Note that $I(\rho)(m) \geq 0$ for any $m \in M$ by the convexity. For any $\alpha > 0$, $\epsilon > 0$ and a sequence of measurable sets A_n with $P(A_n) \to 0$, we have that

$$\begin{aligned} 0 &\leq I(\rho)(-\alpha 1_{A_n}) = \inf_{m\in M} \rho(m - \alpha 1_{A_n}) - \inf_{m\in M} \rho(m) \\ &\leq (1-\epsilon) \inf_{m\in M} \rho\left(\frac{m}{1-\epsilon}\right) + \epsilon\rho\left(-\frac{\alpha}{\epsilon} 1_{A_n}\right) - \inf_{m\in M} \rho(m) \\ &\to -\epsilon \inf_{m\in M} \rho(m) \end{aligned} \tag{3.6}$$

as $n \to \infty$ by the Lebesgue property of ρ. Because ϵ is arbitrary, we conclude the Lebesgue property of $I(\rho)$ by Theorem 2.5. □

PROPOSITION 3.8. *For a finite convex risk measure* ρ, *the following are equivalent:*

1. ρ *is a GDV with the Lebesgue property.*
2. *there exists a convex risk measure* η *with the Lebesgue property,* $\rho = I(\eta)$.

Proof. 1⇒2: This is because $\rho = I(\rho)$ by Theorem 3.4.
2⇒1: By Lemma 2.10, we have $\inf_{m\in M} \eta(m) \in \mathbb{R}$, and so

$$I(\eta)(x) = \inf_{m\in M} \eta(x+m) - \inf_{m\in M} \eta(m).$$

In particular we have (3.5) and (3.6) with η instead of ρ. By the finiteness of $\rho = I(\eta)$, Theorem 2.5 can be applied to have the result. □

EXAMPLE 3.9. Here we give an example of risk indifference price $I(\rho)$ which does not have the Fatou property whereas ρ does. Consider $L = L^\infty(\mathbb{R}, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a normal distribution on $\mathbb{R}$. Let $Q \in \mathcal{P}$ have a compact support and define a sequence $\{Q_n\} \subset \mathcal{P}$ by $Q_n(A) := Q(A - n)$ for $A \in \mathcal{F}$, $n \in \mathbb{N}$. Because $\{g \in L^*_+ | g(1) = 1\}$ is weak-* compact, there exists a cluster point μ of $\{Q_n\}$. Because $\{Q_n\}$ is not tight, $\mu \notin \mathcal{P}$. Consider $M = \{x \in L | \mu(x) \leq 0\}$. Observe that $\overline{L}^* = \{\mu\}$. In fact if there exists $\nu \in \overline{L}^*$ and $x \in L$ with $\nu(x) > \mu(x)$, then $y := x - \mu(x) \in M$ and $\nu(y) > 0$, which is a contradiction. Now consider $\rho \in \mathcal{R}$ defined as $\rho(-x) = \sup_{Q\in\mathcal{P}} \mathbb{E}_Q[x]$. Let us show that $\rho^*(\mu) = 0$. By $\rho(0) = 0$ we have $\rho^*(\mu) \geq 0$ and $\rho^*(Q_n) = 0$. If $\rho^*(\mu) > 0$, then there exists $x \in L$ such that $\mu(x) > \sup_{Q\in\mathcal{P}} \mathbb{E}_Q[x]$, which contradicts that μ is a cluster point of Q_n. By the same reason, we have also that for any $m \in M$ and $x \in L$, $\rho(m + x) \geq \mu(-m - x) \geq -\mu(x)$, so that $I(\rho)$ is finite. By Lemma 2.10, $I(\rho)^*(g) = \infty$ for any $g \in L^*_+ \setminus \overline{L}^*$, so by Theorem 2.2, we have $I(\rho)(-x) = \mu(x)$ for all $x \in L$. To see that $I(\rho)$ does not have the Fatou property, consider the increasing sequence $x_n := 1_{(-\infty,n)}$. Then $I(\rho)(-x_n) = 0$ whereas $I(\rho)(-x_\infty) = 1$.

3.3. Shortfall Risk Measures

Here we treat shortfall risk measure as an application. We presume an investor who sells a claim x. When she sells x with price c and selects $m \in M$ as her strategy, her final cash-flow is $c + m - x$, and so its shortfall is $(c + m - x) \wedge 0$. In general, shortfall risk is defined as a weighted expectation of the shortfall with a loss function. A loss function is a continuous strictly increasing convex function $l : \mathbb{R}_+ \to \mathbb{R}_+$ with $l(0) = 0$.

This represents the seller's attitude toward risk. To suppress the shortfall risk less than a certain level $\delta > 0$ which she can endure, the least price she can accept is given as

$$\rho_l(-x) := \inf\{c \in \mathbb{R} \mid \text{there exists } m \in M \text{ such that } E[l((c+m-x)^-)] \leq \delta\}. \tag{3.7}$$

As shown in Arai (2011) and Föllmer and Schied (2002), ρ_l is a convex risk measure and it has the Fatou property under mild conditions. However, it is not a GDV as $\rho_l(0) \neq 0$:

PROPOSITION 3.10. *Any shortfall risk measure is not a GDV.*

Proof. For any shortfall risk measure ρ_l, (3.7) implies that

$$\begin{aligned}\rho_l(0) &= \inf\{c \in \mathbb{R} \mid \text{there exists } m \in M \text{ such that } E[l((c+m)^-)] \leq \delta\}\\ &\leq \inf\{c \in \mathbb{R} \mid l(c^-) \leq \delta\} = -l^{-1}(\delta) < 0.\end{aligned}$$

Hence, $\rho_l \notin \mathcal{R}$, from which ρ_l is not a GDV. □

Now we show that a normalized shortfall risk measure can be a GDV. Define $\widehat{\rho_l}$ as $\widehat{\rho_l}(x) := \rho_l(x) - \rho_l(0)$.

PROPOSITION 3.11. *If $\widehat{\rho_l} \in \mathcal{R}$, then $\widehat{\rho_l}$ is a GDV.*

Proof. In light of Theorem 3.4, it suffices to see $I(\widehat{\rho_l}) = \widehat{\rho_l}$. Because $\widehat{\rho_l}(m) \geq -\widehat{\rho_l}(-m) \geq 0$ for $m \in M$, we have $\inf_{m \in M} \widehat{\rho_l}(m) = 0$, and so $I(\widehat{\rho_l})(x) = \inf_{m \in M} \rho_l(m + x) - \rho_l(0)$. Now let us observe that $\inf_{m \in M} \rho_l(m + x) = \rho_l(x)$ for any $x \in L$. $\inf_{m \in M} \rho_l(m + x) \leq \rho_l(x)$ holds clearly. Fix $m \in M$ and $c > \rho_l(m + x)$ arbitrarily. Then there exists $m' \in M$ such that $E[l((c + m' + m + x)^-)] \leq \delta$. Because $m' + m \in M$, we have $c \geq \rho_l(x)$. □

4. RELEVANT GOOD DEAL VALUATIONS

4.1. Fundamental Theorem of Asset Pricing

We have seen that the condition $\mathcal{Q} \neq \emptyset$ is equivalent to the existence of a GDV. Example 4.1 below shows that $\mathcal{Q} \neq \emptyset$ is not sufficient to rule out arbitrage opportunities in general.

EXAMPLE 4.1. Let $A \in \mathcal{F}$ with $P(A) \in (0, 1)$, $m' := 1_A$ and $M = \{cm' \mid c \geq 0\} - L_+$. Any probability measure $Q \in \mathcal{P}$ with $Q(A) = 0$ is in $\mathcal{Q}$. On the other hand, cm' with $c > 0$ brings an arbitrage opportunity.

Kreps (1981) showed that $\mathcal{Q}^e \neq \emptyset$ is equivalent to NFL, that is, $\overline{M} \cap L_+ = \{0\}$. Here we prove that $\mathcal{Q}^e \neq \emptyset$ is equivalent to the existence of a relevant GDV, that is, a relevant convex risk measure which is a GDV.

THEOREM 4.2. *The following are equivalent:*

1. $\mathcal{Q}^e \neq \emptyset$.
2. $\overline{M} \cap L_+ = \{0\}$.
3. *There exists a relevant GDV.*

Proof. 2⇒1: For any $a, b \in \mathbb{R}$, the set $\{x \in L | a \leq x \leq b\}$ is compact in $\sigma(L, L^\dagger)$. In fact if $L = L^\infty$, then $L^\dagger = L^1$ and $\sigma(L, L^\dagger)$ is the weak-* topology. The compactness then follows from the Banach-Alaoglu theorem. It suffices then to notice that $L^\infty \subset L$, $L^\dagger \subset L^1$ as sets of random variables and the natural inclusion $(L^\infty, \sigma(L^\infty, L^1)) \to (L, \sigma(L, L^\dagger))$ is continuous. Therefore, we can prove the existence of an element of $\mathcal{Q}^e$ in exactly the same manner as in the proof of Theorem 5.2.3 of Delbaen and Schachermayer (2006).

1⇒3: This is from Lemma 2.7.

3⇒2: Let ρ be a relevant GDV. We have $\rho(x) = \sup_{Q \in \mathcal{Q}}\{\mathbb{E}_Q[-x] - c(Q)\}$ by Item 3 of Theorem 3.4. Because $\rho(-z) > 0$ for all $z \in L_+$ by the relevance, it suffices to see that $\rho(-\overline{m}) \leq 0$ for any $\overline{m} \in \overline{M}$. If there exists $\overline{m} \in \overline{M}$ with $\rho(-\overline{m}) > 0$, there exists $Q \in \mathcal{Q}$ such that $\mathbb{E}_Q[\overline{m}] > c(Q) \geq \sup_{m \in M} \mathbb{E}_Q[m]$. The last inequality is from the fact that $\rho(-m) \leq 0$ for all $m \in M$. This contradicts that $\overline{m}$ is in the closure of M in $\sigma(L, L^\dagger)$. □

Now we give a set of equivalent conditions for GDV to be relevant. Let

$$\begin{aligned} M^\rho &:= \{x - \rho(-x) | x \in L, \rho(-x) < \infty\} - L_+ = \{x \in L | \rho(-x) = 0\} - L_+ \\ &= \{x \in L | \rho(-x) \leq 0\}. \end{aligned} \tag{4.1}$$

Note that M^ρ is a convex set including M and interpreted as the set of the 0-attainable claims of an extended market where an investor offers prices for all $x \in L$ by using ρ as her pricing functional. In light of Theorem 2.3, M^ρ is closed in $\sigma(L, L^\dagger)$. Therefore NFL for this extended market is $M^\rho \cap L_+ = \{0\}$.

THEOREM 4.3. *For a GDV ρ, the following are equivalent:*

1. *ρ is relevant.*
2. *$-\widehat{\rho^0}(x - z) < \rho(-x)$ for any $x \in L$ and $z \in L_+ \backslash \{0\}$.*
3. *$-\rho^0(x - z) < \rho(-x)$ for any $x \in L$ and $z \in L_+ \backslash \{0\}$.*
4. *$M^\rho \cap L_+ = \{0\}$.*

Proof. 1⇒2: By the relevance and Theorem 3.4, for any $z \in L_+ \backslash \{0\}$, there exists $Q(z) \in \mathcal{Q}$ such that $\mathbb{E}_{Q(z)}[z] > \rho^*(Q(z))$. Therefore,

$$-\widehat{\rho^0}(x - z) = \inf_{Q \in \mathcal{Q}} \mathbb{E}[x - z] \leq \mathbb{E}_{Q(z)}[x - z] < \mathbb{E}_{Q(z)}[x] - \rho^*(Q(z)) \leq \rho(-x).$$

2⇒3: This is from Lemma 2.7.

3⇒1: For a given $z \in L_+ \backslash \{0\}$, let $x = z$.

1⇒4: This is because ρ separates M^ρ and $L_+ \backslash \{0\}$.

4⇒1: If ρ is not relevant, then there exists $z \in L_+ \backslash \{0\}$ such that $\rho(-z) = 0$. In particular $z \in M^\rho$, which is a contradiction. □

REMARK 4.4. Item 3 of Theorem 4.3 is the no-near-arbitrage condition (NNA) introduced in Staum (2004). Theorem 6.2 of Staum (2004) states a condition under which ρ_A satisfies NNA. Proposition 4.5 below may be regarded as its counterpart.

PROPOSITION 4.5. *Let ρ be a GDV. If there exists $Q_0 \in \mathcal{Q}^e$ such that $\rho^*(Q_0) = 0$, then ρ is relevant. The reverse implication holds true if ρ is coherent.*

Proof. The relevance is clear from Theorem 2.3. The converse is the Halmos-Savage theorem (see, e.g., Delbaen 2002). □

Note that for $Q \in \mathcal{P}$ and $\rho \in \mathcal{R}$, $\rho^*(Q) = 0$ is equivalent to that $-\rho(x) \leq \mathbb{E}_Q[x] \leq \rho(-x)$ for all $x \in L$. Therefore such Q is interpreted as a consistent pricing kernel of the extended market M^ρ. The following example shows that the coherence in the second assertion of Proposition 4.5 cannot be dropped. In other words, there is no strictly positive consistent pricing kernel in general even if M^ρ satisfies NFL: $M^\rho \cap L_+ = \{0\}$.

EXAMPLE 4.6. We illustrate, by using a binomial model, a noncoherent relevant GDV ρ such that there is no $Q_0 \in \mathcal{Q}^e$ satisfying $\rho^*(Q_0) = 0$. Set $\Omega = \{\omega_1, \omega_2\}$, and $M = L_-$. Denoting $q := Q(\{\omega_1\})$, we can identify q with $Q \in \mathcal{Q}$. From this viewpoint, $\mathcal{Q}$ and $\mathcal{Q}^e$ are corresponding to [0, 1] and (0, 1) respectively. Consider $\rho(-x) = \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[x] - c(Q)\}$ with $c(Q) = q^2$. Then we have $\rho^*(Q) = c(Q)$. Denoting $z_i := z(\omega_i)$ for $i = 1, 2$, we have $\rho(-z) = \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[z] - c(Q)\} = \sup_{q\in[0,1]}\{qz_1 + (1-q)z_2 - q^2\} = \sup_{q\in(0,1)}\{qz_1 + (1-q)z_2 - q^2\} > 0$ for any $z \in L_+\backslash\{0\}$. Thus, ρ is a noncoherent relevant GDV. On the other hand, there is no $q \in (0, 1)$ with $c(Q) = 0$.

REMARK 4.7. In Bion-Nadal (2009), NFL refers to the condition that

$$\overline{\mathrm{cone}(M^\rho)} \cap L_+ = \{0\},$$

where $\overline{\mathrm{cone}(M^\rho)}$ is the closure of $\mathrm{cone}(M^\rho) = \{\lambda m; m \in M^\rho, \lambda \geq 0\}$ in $\sigma(L, L^\dagger)$, which is a different condition to $M^\rho \cap L_+ = \{0\}$ unless ρ is coherent. This alternative definition of NFL enabled to establish the equivalence between NFL of ρ and the existence of $Q_0 \in \mathcal{Q}^e$ with $\rho^*(Q_0) = 0$ in Bion-Nadal (2009). In fact because $\overline{\mathrm{cone}(M^\rho)}$ becomes a cone, the same argument as the proof of 2⇒1 of Theorem 4.2 can apply to have $Q_0 \in \mathcal{Q}^e$ with $\mathbb{E}_{Q_0}[m] \leq 0$ for all $m \in M^\rho$. Because $x - \rho(-x) \in M^\rho$ for all $x \in L$, we have $\rho^*(Q_0) = 0$. Note however that $\mathrm{cone}(M^\rho)$ does not have any interpretation as the set of the 0-attainable claims in general. For instance, in the model of the preceding example, we can find $x \in L$ with $\rho(-x) \leq 0$ and $\lambda > 0$ satisfying $\rho(-\lambda x) > 0$. Therefore, it seems not adequate, from economical point of view, to adapt such a definition of NFL. Consequently, the existence of Q_0 with $\rho^*(Q_0) = 0$ may not be considered as a necessary condition for ρ to be a reasonable pricing functional.

4.2. When Are All Good Deal Valuations Relevant?

As seen in Theorem 4.3, when we extend the underlying market M to M^ρ by using a GDV ρ as pricing functional, the extended market M^ρ remains to satisfy NFL if and only if ρ is relevant. Therefore markets in which any GDV is relevant are stable against such extensions of the market. Here we study necessary and (or) sufficient conditions under which all (coherent) GDVs are relevant.

THEOREM 4.8. *Suppose $\mathcal{Q}^e \neq \emptyset$ and consider the following conditions:*

1. *Any GDV is relevant.*
2. $\widehat{\rho^0}(z) < 0$ *for any* $z \in L_+\backslash\{0\}$.
3. $\mathcal{Q} = \mathcal{Q}^e$.

3′. $\mathcal{Q} = \mathcal{Q}^e$ *and* $\mathcal{Q}^e$ *is* $\sigma(L^\dagger, L)$*-compact.*

4. *Any coherent GDV is relevant.*

Then, we have 1 $\Leftrightarrow$2, 2$\Rightarrow$3, 3$'$$\Rightarrow$2, 3$\Leftrightarrow$4.

Proof. 1$\Rightarrow$2: Assume that there exists $z_0 \in L_+\backslash\{0\}$ such that $\widehat{\rho^0}(z_0) = 0$. Then $\inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[z_0] = 0$, so that we can define $\rho \in \mathcal{R}$ as

$$\rho(-x) = \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[x] - \mathbb{E}_Q[z_0]\}.$$

This is a GDV by Theorem 3.4 but not relevant. In fact $\rho(-z_0) = 0$.

2$\Rightarrow$1: Let ρ be a GDV. Then by Item 5 of Theorem 3.4, $\rho(-z) \geq -\widehat{\rho^0}(z) > 0$ for any $z \in L_+\backslash\{0\}$.

2$\Rightarrow$3: If $\mathcal{Q} \neq \mathcal{Q}^e$, then there exists $Q^* \in \mathcal{Q}\backslash\mathcal{Q}^e$. Denoting $A = \{\mathrm{d}Q^*/\mathrm{d}\mathbb{P} > 0\}$, $\widehat{\rho^0}(1_{A^c}) = \sup_{Q\in\mathcal{Q}} \mathbb{E}_Q[-1_{A^c}] \geq E_{Q^*}[-1_{A^c}] = 0$, whereas $1_{A^c} \in L_+ \setminus \{0\}$.

3$'$$\Rightarrow$2: By compactness we have for any $z \in L_+\backslash\{0\}$,

$$\widehat{\rho^0}(z) = \sup_{Q\in\mathcal{Q}} \mathbb{E}_Q[-z] = \sup_{Q\in\mathcal{Q}^e} \mathbb{E}_Q[-z] = \max_{Q\in\mathcal{Q}^e} \mathbb{E}_Q[-z] < 0.$$

3$\Rightarrow$4: Any coherent GDV ρ is represented as $\rho(x) = \sup_{Q\in\widehat{\mathcal{Q}}} \mathbb{E}_Q[-x]$, for some convex set $\widehat{\mathcal{Q}} \subset \mathcal{Q} = \mathcal{Q}^e$. Therefore ρ is relevant.

4$\Rightarrow$3: If $\mathcal{Q} \neq \mathcal{Q}^e$ then we can take Q^* and A in the same way as "2$\Rightarrow$3". Let $\rho(x) = \sup_{Q\in\mathcal{Q},\, Q(A)=1} \mathbb{E}_Q[-x]$. Then $\rho \in \mathcal{R}$ because $Q^*(A) = 1$. By Theorem 3.4, ρ is a coherent GDV but not relevant because $\rho(1_{A^c}) = 0$. □

The implications "3$\Rightarrow$3$'$", "3$\Rightarrow$1 (or 2)" and "2$\Rightarrow$3$'$" in Theorem 4.8 do not hold in general. We illustrate counterexamples.

EXAMPLE 4.9. We give an example satisfying Item 3 of Theorem 4.8 which does not satisfy Items 1 nor 3$'$. Set $\Omega = \mathbb{R}$, $L = L^\infty$ and $\mathbb{P}(\mathrm{d}u) = \phi(u)\mathrm{d}u$, where $\phi(u)$ is the standard normal density. We consider the set of the mixed normal distributions. Let V be the set of all probability measures on $(0, \infty)$,

$$Q_\mu(\mathrm{d}u) := \int \frac{1}{\sqrt{v}}\phi(u/\sqrt{v})\mu(\mathrm{d}v)\mathrm{d}u$$

for $\mu \in V$, and $\widehat{\mathcal{Q}} := \{Q_\mu | \mu \in V\}$. Define M as

$$M = \{m \in L^\infty | \mathbb{E}_Q[m] \leq 0 \text{ for any } Q \in \widehat{\mathcal{Q}}\}.$$

Note that all bounded odd functions are in M and $\widehat{\mathcal{Q}} \subset \mathcal{Q}^e \subset \mathcal{Q}$. Now we show that $\widehat{\mathcal{Q}}$ is $\sigma(L^1, L^\infty)$-closed. Let $\{\mu_n\} \subset V$ be a sequence with $Q_{\mu_n} \to Q$ in $\sigma(L^1, L^\infty)$. Denote $y_w(u) := e^{iwu}$ for any $w, u \in \mathbb{R}$, where $i = \sqrt{-1}$. We have

$$E_{Q_{\mu_n}}[y_w] = \int e^{-\frac{v}{2}w^2}\mu_n(dv),$$

which has the form of the Laplace transform of μ_n. Because $E_{Q_{\mu_n}}[y_w] \to \mathbb{E}_Q[y_w]$ and $\lim_{w\to 0} \mathbb{E}_Q[y_w] = 1$, the continuity theorem of Laplace transforms (see theorem XIII.1.2 of Feller 1971) implies the existence of $\mu \in V$ such that

$$\mathbb{E}_Q[y_w] = \int e^{-\frac{v}{2}w^2}\mu(\mathrm{d}v),$$

which is the characteristic function of an element of $\widehat{\mathcal{Q}}$. Hence, $Q \in \widehat{\mathcal{Q}}$.

Note that $\widehat{\mathcal{Q}} = \mathcal{Q}^e = \mathcal{Q}$. In fact if there exists $Q^* \in \mathcal{Q} \setminus \widehat{\mathcal{Q}}$, by the Hahn-Banach theorem there exists $x \in L$ such that $\mathbb{E}_{Q^*}[x] > \sup_{Q\in\hat{\mathcal{Q}}} \mathbb{E}_Q[x] =: \alpha$. However, $x - \alpha \in M$ and $\mathbb{E}_{Q^*}[x-\alpha] > 0$, which contradicts $Q^* \in \mathcal{Q}$. On the other hand, $\mathcal{Q}$ is not compact. In fact for the sequence $\mu_n := \delta_{1/n}$ for $n \in \mathbb{N}$, where δ_u is the Delta measure concentrated on $\{u\}$, $\{Q_{\mu_n}\}$ does not have a cluster point in $\widehat{\mathcal{Q}}$.

Finally, we construct a GDV ρ which is not relevant. Letting $y(u) := u^2$, we define ρ as $\rho(-x) = \sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[x] - c(Q)\}$ with $c(Q) = \mathbb{E}_Q[y]$. Obviously, we have $\rho(0) = 0$ and $\rho(-y) = 0$.

EXAMPLE 4.10. Here we see that the implication "2⇒3′" in Theorem 4.8 does not hold. We modify Example 4.9 as follows. Let $\mu_0 \in V$ be fixed and $\widehat{\mathcal{Q}}_0 := \{Q_\nu | \nu = (\mu_0 + \mu)/2, \mu \in V\}$. By the same argument as in Example 4.9, we can prove the closedness and noncompactness of $\widehat{\mathcal{Q}}_0$ and that $\widehat{\mathcal{Q}}_0 = \mathcal{Q} = \mathcal{Q}^e$. This model however satisfies Item 2 of Theorem 4.8 because

$$\widehat{\rho^0}(z) = \sup_{Q\in\widehat{\mathcal{Q}}_0} \mathbb{E}_Q[-z] = \frac{1}{2}E_{Q_{\mu_0}}[-z] + \frac{1}{2}\sup_{\mu\in V} E_{Q_\mu}[-z] \le \frac{1}{2}E_{Q_{\mu_0}}[-z] < 0.$$

We conclude the paper with one more example, which is a simple model taking transaction costs into account. In the following example, a model satisfying Item 3′ of Theorem 4.8 is constructed.

EXAMPLE 4.11. We introduce a model with bid-ask spread where all GDVs are relevant. Let $\Omega = \{\omega_0, \omega_1, \ldots, \omega_n\}$ and the Arrow-Debreu securities for the n states $\omega_1, \ldots \omega_n$ be tradable in a market subject to bid-ask spread. Denote by $a_{1,j}, a_{-1,j}$ the ask and bid prices for the state ω_j respectively for each $j = 1, \ldots, n$. Let $D := \{-1, 1\}^n$. If $a_{1,j} \ge a_{-1,j} \ge 0$ for each j and $\sum_j a_{1,j} \le 1$, then for any $d \in D$, a probability measure Q_d on Ω is uniquely determined by $Q_d(\{\omega_j\}) = a_{d(j),j}$ for $j = 1, \ldots, n$, and $Q_d(\{\omega_0\}) = 1 - \sum_{j=1}^n a_{d(j),j}$. Now let

$$M = \{x \in L | \mathbb{E}_d[x] \le 0 \text{ for all } d \in D\} = \left\{x - \max_{d\in D}\mathbb{E}_d[x] | x \in L\right\} - L_+,$$

where $\mathbb{E}_d$ is the expectation under Q_d. Note that any cash-flow $x \in L$ can be uniquely represented as a sum of a constant and the Arrow Debreu securities and that the price for replicating x is $\max_{d\in D}\mathbb{E}_d[x]$. Therefore, M is actually the set of the 0-attainable claims in this market. By the same separation argument as in the preceding examples, we can show

$$\mathcal{Q} = \left\{\sum_{d\in D}\lambda_d Q_d | \lambda_d \ge 0 \text{ for all } d \in D \text{ and } \sum_{d\in D}\lambda_d = 1\right\}.$$

This set is compact because the set of (λ_d) is a finite dimensional simplex. If $a_{-1,j} > 0$ for each j and $\sum_j a_{1,j} < 1$ in addition, then $\mathcal{Q} = \mathcal{Q}^e$ and so, Item 3′ of Theorem 4.8 is satisfied. Consequently, any GDV in this market is relevant. Remark that the additional condition requires market makers not to quote a set of prices which leads an apparent arbitrage opportunity for themselves.

5. CONCLUSIONS

We investigate GDVs thoroughly and introduce some examples throughout this paper. Theorem 3.2 discusses equivalent conditions for the existence of a GDV. The third condition is weaker than the no-arbitrage condition. We can regard the theorem as a new type of FTAP. Next, Theorem 3.4 enumerates equivalent conditions for a given ρ to be a GDV. Among others, we show that $\rho \in \mathcal{R}$ is a GDV if and only if it is a risk indifference price. The "only if" part is our contribution, which indicates that risk indifference price is the only class of meaningful convex risk measures in hedging and pricing theory. Comparison with Staum (2004) is provided in Remark 3.5. In addition, we extend Kreps–Yan theorem in Theorem 4.2, and show the equivalence between NFL and the existence of a relevant GDV. Judging from Theorems 4.2 and 4.3, we can say that a GDV should be relevant to be reasonable. Many open problems which we should solve arise from our results. In particular, we need more results on risk indifference price. What is a sufficient condition other than the Lebesgue property to have the Fatou property? What is a sufficient condition to be relevant? Except for risk indifference price, extensions of our results to a dynamic version must be significant. It seems that we can easily extend our results, say, Theorem 3.4 to a dynamic version. On the other hand, we need to pay attention to the time-consistency in the dynamic case. The time-consistency of GDVs might bring many complicated problems to us.

REFERENCES

ARAI, T. (2011): Good Deal Bounds Induced by Shortfall Risk, *SIAM J. Financial Math.* 2, 1–21.

BECHERER, D. (2009): From Bounds on Optimal Growth Toward a Theory of Good-deal Hedging, in *Advanced Financial Modelling*, Radon Series on Computational and Applied Mathematics, Vol. 8, Berlin: Walter de Gruyter, pp. 27–51.

BERNARDO, A. E., and O. LEDOIT (2000): Gain, Loss, and Asset Pricing, *J. Pol. Econ.* 108, 144–172.

BIAGINI, S., and M. FRITTELLI (2004): On the Super Replication Price of Unbounded Claims, *Ann. Appl. Prob.* 14, 1970–1991.

BIAGINI, S., and M. FRITTELLI (2009): On the Extension of the Namioka-Klee Theorem and on the Fatou Property for Risk Measures, in *Optimality and Risk: Modern Trends in Mathematical Finance*, *The Kabanov Festschrift*, F. Delbaen, M. Rasonyi, and C. Stricker, eds., Berlin: Springer, pp. 1–28.

BIAGINI, S., M. FRITTELLI, and M. GRASSELLI (2011): Indifference Price with General Semimartingales, *Math. Finance* 21, 423–446.

BION-NADAL, J. (2009): Bid-Ask Dynamic Pricing in Financial Markets with Transaction Costs and Liquidity Risk, *J. Math. Econ.* 45, 738–750.

BJÖRK, T., and I. SLINKO (2006): Towards a General Theory of Good-Deal Bounds, *Rev. Finance* 10, 221–260.

CARR, P., H. GEMAN, and D. B. MADAN (2001): Pricing and Hedging in Incomplete Markets, *J. Finan. Econ.* 62, 131–167.

ČERNÝ, A., and S. HODGES (2002): The Theory of Good-Deal Pricing in Financial Markets, in *Mathematical Finance—Bachelier Congress 2000*, H. Geman, D. Madan, S. R. Pliska, and T. Vorst, eds., Berlin: Springer, pp. 175–202.

ČERNÝ, A. (2003): Generalized Sharpe Ratios and Asset Pricing in Incomplete Markets, *European Finance Review* 7, 191–233.

CHERIDITO, P., and T. LI (2009): Risk Measures on Orlicz Hearts, *Math. Finance* 19, 189–214.

COCHRANE, J. H., and J. SAÁ-REQUEJO (2000): Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets, *J. Pol. Econ.* 108, 79–119.

DELBAEN, F. (2002): *Coherent Risk Measures on General Probability Spaces*, Advances in Finance and Stochastics, Berlin: Springer, pp. 1–37.

DELBAEN, F., and W. SCHACHERMAYER (2006): *The Mathematics of Arbitrage*, Springer Finance. Berlin: Springer-Verlag.

ELLIOTT, R. J., and T. K. SIU (2010): Risk-Based Indifference Pricing under a Stochastic Volatility Model, *Comm. Stoc. Anal.* 4, 51–73.

FELLER, W. (1971): *An Introduction to Probability Theory and Its Applications* Vol. II. 2nd ed., New York-London-Sydney: John Wiley & Sons, Inc.

FÖLLMER, H., and A. SCHIED (2002): Convex Measures of Risk and Trading Constraints, *Finance Stoch.* 6, 429–447.

JASCHKE, S., and U. KÜCHLER (2001): Coherent Risk Measures and Good-Deal Bounds, *Finance Stoch.* 5, 181–200.

JOUINI, E., W. SCHACHERMAYER, and N. TOUZI (2006): Law Invariant Risk Measures Have the Fatou Property, *Adv. Math. Econ.* 9, 49–71.

KLÖPPEL, S., and M. SCHWEIZER (2007): Dynamic Utility-Based Good Deal Bounds, *Statist. Decisions* 25, 285–309.

KLÖPPEL, S., and M. SCHWEIZER (2007): Dynamic Indifference Valuation via Convex Risk Measures, *Math. Finance* 17, 599–627.

KREPS, D. M. (1981): Arbitrage and Equilibrium in Economies with Infinitely Many Commodities, *J. Math. Econ.* 8, 15–35.

LARSEN, K., T. PIRVU, S. SHREVE, and R. TÜTÜNCÜ (2005): Satisfying Convex Risk Limits by Trading, *Finance Stoch.* 9, 177–195.

ØKSENDAL, B., and A. SULEM (2009): Risk Indifference Pricing in Jump Diffusion Markets, *Math. Finance* 19, 619–637.

RAO, M. M., and Z. D. REN (1991): *Theory of Orlicz Spaces*, New York: Marcel Dekker, Inc.

STAUM, J. (2004): Fundamental Theorems of Asset Pricing for Good Deal Bounds, *Math. Finance* 14, 141–161.

XU, M. (2006): Risk Measure Pricing and Hedging in Incomplete Markets, *Ann. Finance* 2, 51–71.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 485–504

THE TWO FUNDAMENTAL THEOREMS OF ASSET PRICING FOR A CLASS OF CONTINUOUS-TIME FINANCIAL MARKETS

ANDREW LYASOFF

Boston University

The paper is concerned with the first and the second fundamental theorems of asset pricing in the case of nonexploding financial markets, in which the excess-returns from risky securities represent continuous semimartingales with absolutely continuous predictable characteristics. For such markets, the notions of "arbitrage" and "completeness" are characterized as properties of the distribution law of the excess-returns. It is shown that any form of arbitrage is tantamount to guaranteed arbitrage, which leads to a somewhat stronger version of the first fundamental theorem. New proofs of the first and the second fundamental theorems, which rely exclusively on methods from stochastic analysis, are established.

KEY WORDS: arbitrage and completeness of financial markets, the first and the second fundamental theorems of asset pricing, Itô processes, predictable representation of local martingales, extremal martingale measures.

1. PRELIMINARIES AND INTRODUCTION

The main objective of this paper is to establish conditions for arbitrage and completeness in continuous-time financial markets that can be formulated entirely in terms of the statistical properties of directly observable market data. We shall restrict our study to market models in which the aggregate excess-returns process associated with any given asset follows a continuous semimartingale with absolutely continuous (for the Lebesgue measure) predictable characteristics.[1] We insist—and this point is important—that the local martingale part and the bounded variation part in the excess-returns may not be directly observable.

The second objective of the paper is to connect certain properties of the distribution law of the excess-returns with the classical formulation of the first and the second fundamental theorems of asset pricing (FTAP), and obtain new proofs to these two theorems, which rely entirely on classical methods from stochastic analysis; specifically, on the methods developed in Liptser and Shiryaev (1974), ch. 7. One consequence of this approach (and choice of market model) is that it allows for a somewhat stronger version of the first fundamental theorem: the existence of arbitrage entails the existence of guaranteed arbitrage, so that the first theorem actually provides necessary and sufficient conditions for guaranteed arbitrage—see Proposition 3.1. This feature illustrates

Manuscript received July 2010; final revision received October 2011
Address correspondence to Andrew Lyasoff, Boston University School of Management, 595 Commonwealth Ave, Boston, MA 02215; e-mail: alyasoff@bu.edu.

[1] While the class of such semimartingales is identical to the general class of Itô-processes, we shall not impose any concrete Itô-structure, in that no "Brownian motion" or "volatility matrix" will be specified exogenously. However, we shall work with filtrations and Itô-structures that are obtained *endogenously* from the semimartingale structure of the excess-returns.

DOI: 10.1111/j.1467-9965.2012.00530.x

the profound difference in nature between the notion of arbitrage in continuous time setting and in discrete time setting.

The connection between the notion of arbitrage and the martingale property of the excess-returns was first established in Harrison and Kreps (1979) and Harrison and Pliska (1981), while the term "fundamental theorem of asset pricing" was coined in Dybvig and Ross (1987). In its full generality the FTAP was established in the groundbreaking papers Delbaen and Schachermayer (1994, 1998). In spite of the many publications that followed the work of F. Delbaen and W. Schachermayer (see Kabanov and Kramkov 1994; Levental and Skorohod 1995; Kabanov and Stricker 2001, for example), the proof of the FTAP in a reasonably general continuous-time setting has resisted a major simplification. To some degree, this paper grew out of the desire to find a simpler proof and formulation of the FTAP and, hopefully, a new understanding of the nature of this important statement. For a detailed account of the background and the history of the FTAP, as well as a comprehensive list of references, we refer to the monograph Delbaen and Schachermayer (2010).

We begin with the general description of the market model this paper is concerned with. The model involves a risk-free rate ($r_t \in \mathbb{R}$) and a finite number $n \geqslant 1$ of risky securities, which will be identified by their respective share prices ($S_t \in \mathbb{R}^n$). For brevity, we shall not suppose that the risky securities pay dividends—nothing will change if the model includes dividends, as long as the semimartingale representation for the aggregate excess-returns stands (see equation (1.1)). There is a finite time horizon $T > 0$, and $(r_t)_{t\in[0,T]}$ and $(S_t)_{t\in[0,T]}$ are treated as adapted stochastic processes[2] on some (sufficiently rich) filtered structure $(\Omega, (\mathscr{F}_t)_{t\in[0,T]}, P)$, the filtration $(\mathscr{F}_t)$ being right-continuous and P-complete (i.e., $\mathscr{F}_0$ contains all sets $A \in \mathscr{F}_T$ with $P(A) = 0$).[3] All components of the price vector (S_t) are assumed to be *continuous exponential* $(\mathscr{F}_t, P)$*-semimartingales* and the risk-free rate (r_t) is assumed to satisfy the usual integrability condition:

$$\int_0^T |r_s| \, \mathrm{d}s < \infty, \qquad P\text{-a.e. in } \Omega\,.$$

For the sake of better readability, throughout the main part of the paper we shall be concerned only with the one-dimensional case ($n = 1$). Except for the choice of the square root, which is discussed in Remark 1.2, the multidimensional case ($n \geqslant 2$) essentially differs only in the notation—to quote from a well-known paper on the subject: "If the reader is willing to accept the 1-dimensional notation for the n-dimensional case as well, nothing has to be changed." Nevertheless, the nature of the method that we are about to develop is such that keeping in mind the general case ($n \geqslant 2$) still provides valuable insight and intuition—see footnotes 5 and 6, the comments following (1.5), and Remark 1.5.

[2] Throughout the present exposition any use of the term "process" automatically implies that the respective object is measurable as a function on the product space $\Omega \times [0, T]$, relative to the product σ-algebra $\mathscr{F}_T \otimes \mathscr{B}([0, T])$.

[3] Unlike the right-continuity of $(\mathscr{F}_t)$, the assumption that $(\mathscr{F}_t)$ is P-complete is not really needed and is included here mostly as a compliance with the tradition. One can avoid the requirement for the underlying filtration to be P-complete, if, instead of working with processes that are continuous P-a.e., one works with processes that are continuous only on some random interval $[0, \tau[$ and vanish on $[\tau, \infty[\cap 1_{\tau<\infty}$, for some $(\mathscr{F}_t)$-stopping time τ with $P[\tau = \infty] = 1$—see Lyasoff (2007) for further explanation. While the completion of the filtration is entirely routine and innocuous in the general theory of stochastic processes, its use in financial context raises the question: what does it mean for a market agent to be informed about events that will occur in the future with probability zero? Addressing such issues is beyond the scope of this paper and, for that reason, we shall work with complete filtrations, although this technical condition can be removed.

With the above-mentioned assumptions and conventions in mind, the *aggregate excess-returns process*, namely

$$\tilde{X}_t := \int_0^t \left(\frac{\mathrm{d}S_v}{S_v} - r_v \,\mathrm{d}v \right), \quad t \in [0, T],$$

can be treated as a continuous $(\mathscr{F}_t, P)$-semimartingale, i.e., it can be expressed as

$$\tilde{X}_t = M_t + A_t, \quad t \in [0, T], \tag{1.1}$$

where (M_t) is some continuous $(\mathscr{F}_t, P)$-local martingale and (A_t) is some continuous and $(\mathscr{F}_t)$-adapted process of finite variation. We now impose the following condition:

ASSUMPTION 1.1. *The measure* $\mathrm{d}A_t$ *is absolutely continuous with respect to the Lebesgue measure* $\mathrm{d}t$ *on* $[0, T]$ *(i.e.,* $\mathrm{d}A_t \prec \mathrm{d}t$*) and the measure* $\mathrm{d}\langle \tilde{X}, \tilde{X}\rangle_t \equiv \mathrm{d}\langle M, M\rangle_t$ *is equivalent to the Lebesgue measure (i.e.,* $\mathrm{d}\langle M, M\rangle_t \sim \mathrm{d}t$*), P-a.e. in* Ω.

We can now set

$$\gamma_t := \frac{\mathrm{d}}{\mathrm{d}t}\langle \tilde{X}, \tilde{X}\rangle_t, \quad \alpha_t := \gamma_t^{\,1/2}, \quad \theta_t := \alpha_t^{\,-1}\frac{\mathrm{d}}{\mathrm{d}t}A_t, \quad B_t := \int_0^t \alpha_s^{\,-1}\,\mathrm{d}M_s,$$

and, as a result, write

$$\tilde{X}_t = \int_0^t \alpha_s \,(\mathrm{d}B_s + \theta_s \,\mathrm{d}s), \quad t \in [0, T]. \tag{1.2}$$

Remarkably, (B_t) is a $(\mathscr{F}_t, P)$-Brownian motion and the first integral above can be treated as a standard Itô-integral. One may also recognize that—formally, at least—the process (θ_t) looks very much as the usual "market price of risk," but it remains to be clarified of *which* risk—see later.

REMARK 1.2. As there are many choices for the square root $\alpha_t = \gamma_t^{\,1/2}$, there are many choices for the representation (1.2), too. Indeed, even if $\gamma_t^{\,1/2}$ is understood to be the positive root, by choosing (ψ_t) to be some—i.e., *any*—$(\mathscr{F}_t)$-adapted stochastic process on Ω with $|\psi_t| = 1$ and setting $\alpha_t := \psi_t\gamma_t^{\,1/2}$ one would still arrive at (1.2), albeit with a different Brownian motion, namely, with the Brownian motion $(\int_0^t \psi_s^{\,-1}\,\mathrm{d}B_s)$, where (B_t) is the Brownian motion constructed with $\alpha_t = \gamma_t^{\,1/2}$. We emphasize—and this point will be very important in the next section—that the square root (α_t) can always be chosen to be adapted to the natural filtration of $(\tilde{X}_t)$. Accordingly, if $\alpha_t := \psi_t\gamma_t^{\,1/2}$, then the process (θ_t) in (1.2) will have to be replaced by $(\psi_t^{\,-1}\,\theta_t)$, and we note that the quantity $(\int_0^T \theta_s^{\,2}\,\mathrm{d}s)$ remains invariant under any such change of the square root (α_t).

The nature of the multiplicity of the square root (α_t) is easier to illustrate in the multidimensional case ($n \geqslant 2$). In this case (γ_t) is a symmetric-matrix-valued process of dimension $n \times n$, with entries $\gamma_t^{i,j} = \frac{\mathrm{d}}{\mathrm{d}t}\langle M^i, M^j\rangle_t \equiv \frac{\mathrm{d}}{\mathrm{d}t}\langle \tilde{X}^i, \tilde{X}^j\rangle_t$, $1 \leqslant i, j \leqslant n$. Furthermore, in the multidimensional version of Assumption 1.1 one must require that γ_t is $\mathrm{d}P \otimes \mathrm{d}t$-a.e. strictly positive definite. By the argument—and also the notation—of theorem (3.9) in Revuz and Yor (1999), ch. V, one can write (in what follows the token $\dagger$ stands for the usual transposition of a matrix) $\gamma_t = \beta_t\,\rho_t\,\beta_t\dagger$, where (β_t) and (ρ_t) are matrix-valued processes of dimension $n \times n$, respectively, with values in the space of orthogonal matrices and the space of diagonal matrices. Both (β_t) and (ρ_t) can be chosen to be adapted to the natural filtration of $(\tilde{X}_t)$. The "square root" α_t can now be identified with the matrix $\beta_t \rho_t^{\frac{1}{2}}$, which has inverse $\alpha_t{}^{-1} = \rho_t^{-\frac{1}{2}} \beta_t{}^{\dagger}$. The argument in the aforementioned result from Revuz and Yor (1999) shows that the process

$$B_t := \int_0^t \alpha_s{}^{-1}\, \mathrm{d} M_s, \qquad t \in [0, T],$$

is an n-dimensional Brownian motion. The relation (1.2) can now be understood in vector form (the vector-stochastic integral being understood exactly as in Revuz and Yor 1999): α_s is a $n \times n$-matrix and B_s and θ_s are vector-columns of dimension n. The choice of the matrix-valued processes (β_t) and (ρ_t) is not unique, and, of course, if $n \geqslant 2$ there is no choice for the matrix-square root that may be seen as "standard," in the same way in which the positive root in the case $n = 1$ is usually understood. Furthermore, in its vector-form, the relation (1.2) is "rotation invariant," in that, if (1.2) holds for some choice of the matrix-process (α_t), the vector-process (θ_t), and the vector-Brownian motion (B_t), then, given *any* orthogonal-matrix-valued stochastic process (O_t), which is adapted to $(\mathscr{F}_t)$—or, in particular, to the natural filtration of $(\tilde{X}_t)$—one can write:

$$\tilde{X}_t = \int_0^t \alpha'_s \left(\mathrm{d} B'_s + \theta'_s \,\mathrm{d}s \right), \qquad t \in [0, T],$$

where

$$\alpha'_t = \alpha_t\, O_t{}^{\dagger}, \qquad B'_t = \int_0^t O_s \,\mathrm{d} B_s, \qquad \theta'_t = O_t\, \theta_t,$$

and, obviously ($\|\cdot\|$ stands for the usual Euclidean norm in $\mathbb{R}^n$),

$$\int_0^T \|\theta_s\|^2 \,\mathrm{d}s = \int_0^T \|\theta'_s\|^2 \,\mathrm{d}s.$$

The following condition will play a crucial role in the rest of the paper:

$$\int_0^T \theta_s^2 \,\mathrm{d}s < \infty, \qquad P\text{-a.e. in } \Omega\,. \tag{1.3}$$

As was shown in Levental and Skorohod (1995), the failure of the above condition implies the existence of arbitrage at some moment before the terminal date T.[4] However, our reason to insist that this last condition must hold in any realistic model of financial markets is somewhat different and more primitive: as we are about to demonstrate, condition (1.3) must hold if one is to suppose that the total size of the economy cannot explode to $+\infty$ in finite time with positive probability. There is also an intrinsic reason as to why financial markets should be expected to "control" the size of the market price of risk: excessively large expected instantaneous net returns from risky securities entail excessively large demands for money (to invest in such securities), which, in turn, means higher and higher interest rates, which, in turn, means lower and lower market price of risk. In any case, the precise reason for restricting our study only to market models in which condition (1.3) holds is contained in the following result, the proof of which is given in the Appendix.

[4] What was proved in Levental and Skorohod (1995) is that the violation of (1.3) leads to a stronger form of arbitrage. Arbitrage of that form, referred to as "immediate arbitrage," was studied also in Delbaen and Schachermayer (1995).

PROPOSITION 1.3. *Suppose that condition (1.3) does not hold and that, therefore,*

$$\varepsilon := P\left[\int_0^T {\theta_s}^2\,\mathrm{d}s = \infty\right] > 0\,.$$

Then one can find a constant $0 > C > -\infty$, *possibly depending on* ε, *for which the following claim can be made: given any—arbitrarily large—real number* $\Phi > 0$, *one can find some predictable and* $(\tilde{X}_t)$*-integrable process* (h_t), *possibly depending on* Φ, *which has the following two properties:*

$$P\left[\inf\nolimits_{t\in[0,T]}\int_0^t h_s\,\mathrm{d}\tilde{X}_s \geqslant C\right] = 1 \quad \textit{and} \quad P\left[\int_0^T h_s\,\mathrm{d}\tilde{X}_s \geqslant \Phi\right] \geqslant {}^{\varepsilon}/3\,.$$

One interesting consequence of the last proposition is that eliminating the possibility of "doubling strategies" by imposing "limited liability" on agents' positions does not automatically preclude the possibility of achieving arbitrarily large payoffs with strictly positive probability, which is bounded away from 0. If such a possibility is to be eliminated from the model—as it should, if the total size of the economy cannot explode to $+\infty$ in finite time—then (1.3) must be required.

Since we have not yet made any concrete assumptions about the flow of information that market agents are allowed to observe, the process (h_t) that appears in Proposition 1.3 may not be a trading strategy in the usual sense, in that it may not be possible to construct (h_t) from available information by following a particular strategy. However, we would like to exclude the possibility—in any state of nature whatsoever—that a market agent can attain arbitrarily large wealth (assets less liabilities over time) with strictly positive probability, regardless of whether the agent follows a rational set of rules or simply happens to be lucky. To wit, the event that a market agent attains wealth that is twice the size of the entire world economy should have zero probability in any realistic model, regardless of what information is available to the agent.

As long as condition (1.3) is in force, one can define the local martingale

$$\Theta_t := \int_0^t \theta_s\,\mathrm{d}B_s, \quad t \in [0, T],$$

and then rewrite (1.1) in the following equivalent form:

$$\tilde{X}_t = M_t + \langle M, \Theta\rangle_t\,, \quad t \in [0, T]. \tag{1.4}$$

As we are about to demonstrate (see the next section), writing the excess-returns in the above form is much more useful than the general representation (1.1). The only reason why our starting point was (1.1) and not (1.4) was that (1.4) became possible only after the nonexplosion condition (1.3) was justified in Proposition 1.3. In addition, (1.4) allows for the following interpretation: the local martingale (M_t) represents *the risk* in the financial market, while the local martingale (Θ_t) represents *the pricing rule* for any marketable risk, so that the bracket $\langle M, \Theta\rangle_t$ gives the market price of the aggregate risk $M_t = \int_0^t \mathrm{d}M_s$.[5]

[5] In the multidimensional case ($n \geqslant 2$) the pricing rule for marketable risk is still a scalar-valued local martingale, written as the vector-stochastic integral $\Theta_t := \int_0^t {\theta_s}^\dagger\,\mathrm{d}B_s$, while the bracket $\langle M, \Theta\rangle_t$ is understood in the obvious way as the vector $(\langle M^1, \Theta\rangle_t, \ldots \langle M^n, \Theta\rangle_t)$.

REMARK 1.4. Instead of the usual market price of risk, we prefer to work with what we call "the pricing rule for marketable risk," which is nothing but the Itô-integral of a process that is analogous to the "market price of risk." This slight variation in the terminology is deliberate: we would like to model the pricing operation for any marketable risk (written as a local martingale) as the bracketing operation against the pricing rule (also written as a local martingale). Note that the risk that is being priced in (1.4) is the risk incorporated in the local martingale (M_t)—not the risk incorporated in some Brownian motion from a particular representation of (M_t) in the form of an Itô-integral.

In any case, we can now recast our model in terms of the $(\mathscr{F}_t, P)$-local martingales (M_t) and (Θ_t)—respectively, the market risk and the pricing rule for marketable risk. These two local martingales are required to be continuous, to start from 0 at $t = 0$ and to satisfy (P-a.e.) the conditions: $\langle\Theta, \Theta\rangle_T < +\infty$, $\mathrm{d}\langle\Theta, \Theta\rangle_t \prec \mathrm{d}t$ and $\mathrm{d}\langle M, M\rangle_t \sim \mathrm{d}t$ (i.e., the market risk is assumed to be nondegenerate at all times). The market excess-returns are then given by (1.4). We stress that this description of the market does not involve any concrete representations of the local martingales (M_t) and (Θ_t) in the form of Itô-integrals, even if the existence of such representation is assumed—by theorem (3.9) in Revuz and Yor (1999), ch. V, it follows from the conditions $\mathrm{d}\langle M, M\rangle_t \sim \mathrm{d}t$ and $\mathrm{d}\langle\Theta, \Theta\rangle_t \prec \mathrm{d}t$.

The only component still missing from our description of the financial market is the flow of information that is available to market participants. It is quite common in the finance literature to express the model (1.1) in the Itô-form:

$$\tilde{X}_t = M_t + A_t = \int_0^t \sigma_s \,\mathrm{d}W_s + \int_0^t a_s \,\mathrm{d}s, \quad t \in [0, T], \tag{1.5}$$

with the usual meaning of the symbols involved: $(\sigma_t \in \mathbb{R}^{n\otimes m})$ is a matrix-valued stochastic process, $(a_t \in \mathbb{R}^n)$ is a vector-valued stochastic process, and $(W_t \in \mathbb{R}^m)$ is a Brownian motion of dimension m starting from 0. All these objects are assumed to be defined on the underlying structure $(\Omega, (\mathscr{F}_t)_{t\in[0,T]}, P)$ and to satisfy the minimal integrability and measurability conditions that are needed in order to make the above relation meaningful: (W_t) is a $(\mathscr{F}_t, P)$-Brownian motion and (σ_t) and (a_t) are both adapted to $(\mathscr{F}_t)$ and satisfy

$$\int_0^T \mathrm{Trace}\big[\sigma_s\sigma_s^\dagger\big]\,\mathrm{d}s < \infty \quad \text{and} \quad \int_0^T \|a_s\|\,\mathrm{d}s < \infty, \quad P\text{-a.e. in } \Omega.$$

It is usually assumed that (W_t) is adapted to the information filtration, i.e., all market agents can observe the sample path $t \rightsquigarrow W_t$; even more, in many cases the information filtration is simply taken to be the natural filtration of (W_t). For example, this feature may come as a consequence of the additional requirement that the sample path $t \rightsquigarrow a_t$ is observable, in which case one can always construct—but not in a unique way—a square-matrix-process (σ_t) and a Brownian motion (W_t), both of which are observable, so that (1.5) holds (naturally, the excess-returns $(\tilde{X}_t)$ are always observable). Unfortunately, the requirement that $t \rightsquigarrow a_t$ is observable is hard to justify from a practical point of view and without this requirement there would be no meaningful way in which one can extract information about the matrix (σ_t) and the Brownian motion (W_t) from practical observation. Of course, the process $\sigma_t\sigma_t^\dagger = \gamma_t \equiv \frac{\mathrm{d}}{\mathrm{d}t}\langle\tilde{X}, \tilde{X}\rangle_t$ is always observable, but, generally, information about (γ_t) would not be enough to even determine the dimension of (W_t). In what follows we shall acknowledge that market agents may be receiving information from sources other than the quoted asset prices, but shall not assume any exogenously specified Itô-structure of the local martingales (M_t) and (Θ_t). Even more, we shall *not* impose the requirement that these two local martingales are observable.

REMARK 1.5. Even if the Itô-structure of (1.5) is assumed, in general, the Brownian motion (B_t) in (1.2) may be different from the Brownian motion (W_t) in (1.5) (and (α_t) may be different from (σ_t)). For example, in the multidimensional case—see Remark 1.2—(B_t) is always of the same dimension (n) as the excess-returns $(\tilde{X}_t)$, whereas the dimension of (W_t) may be higher if (σ_t) is rectangular. Thus, if the method developed in this paper is to be applied to the Itô-model in (1.5), in general, one may still need to construct the Brownian motion (B_t) from the local martingale in (1.5)—and rewrite (1.5) in the form (1.2), in spite of the fact that the local martingale in (1.5) is already written as an Itô-integral.

To be precise, we shall assume that all market agents can observe only: (1) the aggregate excess-returns $(\tilde{X}_t)_{t\in[0,T]}$, as defined in (1.4); and (2) an information process $(\bar{X}_t)_{t\in[0,T]}$, which is some continuous and $(\mathscr{F}_t)$-adapted scalar-valued process on $(\Omega, (\mathscr{F}_t)_{t\in[0,T]}, P)$ that starts from 0 (nothing would change if we replace the starting point with any other constant $c \in \mathbb{R}$). In general, just as $(\tilde{X}_t)$, the information process $(\bar{X}_t)$, too, could be multidimensional—our main results are stated for scalar-valued $(\tilde{X}_t)$ and $(\bar{X}_t)$ merely for the sake of greater clarity. The observation process, which we shall also call "the market process," is then the two-dimensional process

$$X_t = (\tilde{X}_t, \bar{X}_t), \quad t \in [0, T],$$

and the observation filtration, which we shall also call "the market filtration," is the natural filtration (on Ω) associated with (X_t). Note that, in general, the market filtration may be smaller than $(\mathscr{F}_t)$, but we do not exclude the case where it is not; for example, the information process $(\bar{X}_t)$ may be the Brownian motion (W_t) in the model (1.5) and $(\mathscr{F}_t)$ may be the filtration generated by (W_t). In the other extreme, the sample path $t \rightsquigarrow \bar{X}_t$ may be nonrandom (say, equal to the constant 0), in which case $(\bar{X}_t)$ would not carry any additional information.

2. EQUIVALENT MEASURES AND ARBITRAGE

In the financial market described in the previous section the agents' world is reduced to the structure

$$\left(\mathbb{X}, (\mathscr{G}_t)_{t\in[0,T]}, \mu\right),$$

in which $\mathbb{X}$ is the space of continuous functions $\mathscr{C}([0, T]; \mathbb{R}^2)$ (endowed with the usual uniform topology and the associated Borel structure), μ is the distribution law of the market process (X_t) under the measure P, and $(\mathscr{G}_t)$ is the smallest right-continuous filtration in $\mathbb{X}$ to which the canonical coordinate process $\chi_t(x) := x_t$, $t \in [0, T]$, $x \in \mathbb{X}$, happens to be adapted. Just as before, we suppose that $\mathscr{G}_0$ is augmented with all μ-negligible events in $\mathscr{G}_T$—and again remark that this assumption is not really necessary. We also note that, since $X_0 = (0, 0)$, the σ-algebra $\mathscr{G}_0$ is μ-a.e. trivial by definition. The filtration $(\mathscr{G}_t)$ is nothing but a replica of the market filtration that was introduced at the end of the previous section.

The market process (X_t), which was defined on Ω, is now statistically indistinguishable from the coordinate process (χ_t), which was defined on $\mathbb{X}$: a market agent has no way of distinguishing between the sample-path $t \rightsquigarrow \chi_t(x) \equiv x_t$ and the sample path $t \rightsquigarrow \chi_t(X_\cdot(\omega)) \equiv X_t(\omega)$, provided that $x \in \mathbb{X}$ is sampled with probability law μ and $\omega \in \Omega$ is sampled with probability law P. Consequently, there is no ambiguity if we refer to $(\chi_t) \equiv (\tilde{\chi}_t, \bar{\chi}_t)$ as the market process, and interpret its first and second coordinates, resp., $(\tilde{\chi}_t)$ and $(\bar{\chi}_t)$, as the excess-returns process and the information process (we have silently adopted the convention to write $\tilde{V}$ for the first coordinate of any two-dimensional vector object V, and write $\bar{V}$ for the second coordinate).

Not all objects introduced in the previous section can be replicated on the market sample-space $\mathbb{X}$: replication is only possible for objects that are adapted to the market filtration and, in general, the market risk (M_t) and the pricing rule for marketable risk (Θ_t) are not in this category. However, the square root (α_t) is, since, as was pointed out in the previous section, (α_t) can always be chosen to be adapted to the natural filtration of $(\tilde{X}_t)$. We now choose and fix some $(\mathscr{G}_t)$-adapted square root of the Gaussian differential $\frac{\mathrm{d}}{\mathrm{d}t}\langle \tilde{\chi}, \tilde{\chi}\rangle_t$. Without any ambiguity, we again denote this root by (α_t); i.e., (α_t) is some arbitrarily chosen $(\mathscr{G}_t)$-adapted process on $\mathbb{X}$ with the property

$$\langle \tilde{\chi}, \tilde{\chi}\rangle_t = \int_0^t {\alpha_s}^2 \,\mathrm{d}s, \quad \text{for all } \; t \in [0, T], \qquad \mu\text{-a.e. in } \mathbb{X}\,.$$

Although the market risk (M_t) cannot be replicated on $\mathbb{X}$, the nondegeneracy property of (M_t) can: we have ${\alpha_t}^2 > 0$, $\mathrm{d}\mu \otimes \mathrm{d}t$-a.e. in $\mathbb{X}$.[6] The process

$$\xi_t := \int_0^t \alpha_s^{-1}\,\mathrm{d}\tilde{\chi}_s, \qquad t \in [0, T],$$

is a well-defined $(\mathscr{G}_t, \mu)$-semimartingale on $\mathbb{X}$, and P-a.e. in Ω one has

$$\xi_t(X_\cdot) = \int_0^t \alpha_s(X_\cdot)^{-1}\,\mathrm{d}\tilde{X}_s \equiv B_t + \langle B, \Theta\rangle_t, \qquad \text{for all } t \in [0, T],$$

where

$$B_t := \int_0^t \alpha_s(X_\cdot)^{-1}\,\mathrm{d}M_s \text{ is } (\mathscr{F}_t, P) - \text{Brownian motion on } \Omega.$$

The process (ξ_t)—note that (ξ_t) is defined on $\mathbb{X}$, not on Ω—will be referred to as *the normalized excess-returns process* and will play a crucial role in our study. The financial interpretation of this process and the use of the term "normalized" should be clear: (ξ_t) is the integral of the instantaneous returns from a dynamically adjusted and self-financed holding of the risky security, or, in the vector case, from n self-financed portfolios of risky securities, represented by the rows of the matrix process (α_t^{-1}). Furthermore, these portfolios are chosen so that the associated market risk is "normalized," in the sense that it constitutes a Brownian motion. The bracket $\langle B, \Theta\rangle_t$ is nothing but the market price of the risk incorporated in the local martingale (B_t), and we again emphasize that (B_t) and $\langle B, \Theta\rangle_t$ *may not be directly observable*—which is why these objects live on Ω, not on $\mathbb{X}$. Nevertheless, the normalized excess-returns process (ξ_t) lives on $\mathbb{X}$ and is $(\mathscr{G}_t)$-adapted by definition, so that it *is* observable. Furthermore, the market agents cannot distinguish

[6] In the multidimensional case ($n \geqslant 2$) this property becomes: $\det[\alpha_t {\alpha_t}^{\dagger}] > 0$, $\mathrm{d}\mu \otimes \mathrm{d}t$-a.e. in $\mathbb{X}$.

between observing $t \rightsquigarrow \xi_t(x)$, $x \in \mathbb{X}$, and observing $t \rightsquigarrow \xi_t(X_{\cdot}(\omega))$, $\omega \in \Omega$, when $x \in \mathbb{X}$ is sampled with law μ and $\omega \in \Omega$ is sampled with law P.

The following definition is completely standard:

DEFINITION 2.1. A "trading strategy" is any $(\mathscr{G}_t)$-adapted stochastic process $(\pi_t \in \mathbb{R})$, which is defined on $\mathbb{X}$ and is $(\tilde{\chi}_t)$-integrable, in that

$$\int_0^T \pi_s{}^2 \, \mathrm{d}\langle \tilde{\chi}, \tilde{\chi} \rangle_s \equiv \int_0^T \pi_s{}^2 \alpha_s{}^2 \, \mathrm{d}s < \infty, \quad \mu\text{-a.e. } in\ \mathbb{X}.$$

We say that the trading strategy (π_t) is "tame," or "admissible," if one can find a constant $0 \geqslant C > -\infty$ with the property

$$\inf_{t \in [0,T]} \int_0^t \pi_s \, \mathrm{d}\tilde{\chi}_s \geq C, \quad \mu\text{-a.e. } in\ \mathbb{X}.$$

The tame trading strategy (π_t) is said to represent "arbitrage" if

$$\mu\left[\int_0^T \pi_s \, \mathrm{d}\tilde{\chi}_s > 0\right] > 0 \quad \text{and} \quad \mu\left[\int_0^T \pi_s \, \mathrm{d}\tilde{\chi}_s \geqslant 0\right] = 1,$$

and is said to represent a "guaranteed arbitrage" if one can find a constant $0 < \varepsilon < \infty$ with the property

$$\mu\left[\int_0^T \pi_s \, \mathrm{d}\tilde{\chi}_s \geqslant \varepsilon\right] = 1.$$

Since the aggregate returns associated with a particular trading strategy (π_t) follow the process

$$\int_0^t \pi_s \, \mathrm{d}\tilde{\chi}_s \equiv \int_0^t \pi_s \, \alpha_s \, \mathrm{d}\xi_s, \quad t \in [0, T],$$

we see that trading with strategy (π_t) on the excess-returns $(\tilde{\chi}_t)$ is no different from trading with strategy $(\pi'_t \equiv \pi_t \, \alpha_t)$ on the normalized excess-returns (ξ_t) (and trading with strategy (π'_t) on the normalized excess-returns (ξ_t) is no different from trading with strategy $(\pi_t \equiv \pi'_t \, \alpha_t{}^{-1})$ on the actual excess-returns $(\tilde{\chi}_t)$).

In connection with the normalized excess-returns (ξ_t), we now introduce our second canonical structure, namely,

$$\big(\mathbb{W}, (\mathscr{A}_t)_{t \in [0,T]}, \nu\big),$$

in which $\mathbb{W}$ is the space $\mathscr{C}([0, T]; \mathbb{R})$, equipped with the usual Borel structure, ν is the distribution law of the semimartingale (ξ_t) (under the measure μ), and $(\mathscr{A}_t)$ is the smallest right-continuous filtration on $\mathbb{W}$ to which the canonical coordinate process on $\mathbb{W}$ happens to be adapted. The standard Wiener measure on the space $\mathbb{W}$, i.e., the distribution law of one-dimensional Brownian motion that starts from 0, we denote by $\mathscr{W}$. We again suppose that $\mathscr{A}_0$ includes all sets $A \in \mathscr{A}_T$ with $\mathscr{W}(A) = 0$, and again remark that this assumption is not really necessary.

This study rests in a crucial way on the result stated in theorem 7.4 from Liptser and Shiryaev (1974). According to this result, the property $\langle \Theta, \Theta \rangle_T < \infty$, P-a.e., guarantees that $\nu \prec \mathscr{W}$, i.e., in the model that we have adopted, the distribution law of the normalized excess-returns is *always* absolutely continuous relative to the canonical Wiener measure on the space $\mathbb{W}$. The complete proof of this statement, which is merely an adaptation of the proof outlined in Liptser and Shiryaev (1974), is provided in the Appendix—see A.1. In any case, the property $\nu \not\sim \mathscr{W}$— i.e., ν not being equivalent to $\mathscr{W}$—is tantamount to the existence of a set $U \in \mathscr{A}_T$ with $\mathscr{W}(U) > 0$ and $\nu(U) \equiv \mu[\xi_{\boldsymbol{\cdot}} \in U] = 0$.

DEFINITION 2.2. An "equivalent local martingale measure" (or, ELMM, for short) is any probability measure Q on the σ-algebra $\mathscr{G}_T$, which is equivalent to the distribution law, μ, of the market process (X_t), (or, $Q \sim \mu$, for short), and is also such that the excess-returns process $(\tilde{\chi}_t)$ is a $(\mathscr{G}_t, Q)$-local martingale on $\mathbb{X}$. The space of all equivalent local martingale measures Q we denoted by$\mathscr{H}$.

We stress that in our setting an ELMM is a measure defined on the agents' space $\mathbb{X}$—not on the underlying sample space Ω. Since $(\tilde{\chi}_t)$ is a $(\mathscr{G}_t, Q)$-local martingale, and since (ξ_t) can be written as a stochastic integral with respect to $(\tilde{\chi}_t)$ (the construction of which depends only on the equivalence class of μ), we can claim that (ξ_t) is a $(\mathscr{G}_t, Q)$-local martingale, for any $Q \in \mathscr{H}$. As the bracket $\langle \xi, \xi \rangle_t$ depends only on the equivalence class of μ, (ξ_t) is actually a $(\mathscr{G}_t, Q)$-Brownian motion on $\mathbb{X}$, for any $Q \in \mathscr{H}$. Conversely, if the measure $Q \sim \mu$ is such that (ξ_t) is a $(\mathscr{G}_t, Q)$-Brownian motion on $\mathbb{X}$, then Q must be an ELMM, since

$$\tilde{\chi}_t = \int_0^t \alpha_s \, \mathrm{d}\xi_s, \quad \text{for all } \ t \in [0, T], \ \mu\text{-a.e. in } \mathbb{X},$$

with the implication that $(\tilde{\chi}_t)$ must be a $(\mathscr{G}_t, Q)$-local martingale for any such Q.

PROPOSITION 2.3. *The property $\mathscr{H} \neq \emptyset$ is equivalent to the property $\nu \sim \mathscr{W}$.*

Proof. We choose and fix some Borel-measurable mapping $\phi : \mathbb{X} \mapsto \mathbb{W}$, which represents the μ-equivalence class of $\xi_{\boldsymbol{\cdot}}$. By the well-known result from measure theory—see proposition 46.3 in Parthasarathy (1977), for example—one can find a transition probability measure $\Pi(w, A)$, $w \in \mathbb{W}$, $A \in \mathscr{G}_T$, with the property $\Pi(w, \phi^{-1}(\{w\})) = 1$, for ν-a.e. $w \in \mathbb{W}$, and the property

$$\mathrm{E}_\nu\left[\int_{\mathbb{X}} f(x)\, \Pi(w, \, \mathrm{d}x)\right] = \mathrm{E}_\mu[\,f\,], \text{ for any } \quad f \in \mathscr{L}^1\,(\mathbb{X}, \mathscr{G}_T, \mu)\,.$$

In particular, $\mu(A) = \mathrm{E}_\nu[\Pi(w, A)]$, for any $A \in \mathscr{G}_T$, and, as a result, we can claim that $\mu(A) = 0$ if and only if $\nu(A^*_\mu) = 0$, where

$$A^*_\mu := \{w \in \mathbb{W} : \Pi(w, A) > 0\}.$$

Now suppose that $\nu \sim \mathscr{W}$ and observe that, under this assumption, ν-negligibility of A^*_μ is the same as $\mathscr{W}$-negligibility. Consequently, if we define the measure

$$Q(A) := \mathrm{E}_{\mathscr{W}}[\Pi(w, A)], \ A \in \mathscr{G}_T,$$

then, given any $A \in \mathscr{G}_T$, we have

$$Q(A) = 0 \quad \Longleftrightarrow \quad \mathscr{W}(A^*_\mu) = 0 \quad \Longleftrightarrow \quad \nu(A^*_\mu) = 0 \quad \Longleftrightarrow \quad \mu(A) = 0,$$

which implies $Q \sim \mu$. Since $\Pi(w, A)$ is a transition measure, given any $B \in \mathscr{A}_T$, we have $\Pi(w, \{\phi \in B\}) = 1_B(w)$, for ν-a.e. $\Longleftrightarrow \mathscr{W}$-a.e. $w \in \mathbb{W}$. Therefore,

$$Q[\phi \in B] = \mathtt{E}_{\mathscr{W}}[1_B(w)] = \mathscr{W}(B)\,, \qquad \text{for any} \qquad B \in \mathscr{A}_T,$$

with the implication that, under Q, the mapping ϕ, and therefore also $\xi_{\boldsymbol{\cdot}}$, is distributed in $\mathbb{W}$ with law $\mathscr{W}$. In other words, (ξ_t) is a $(\mathscr{G}_t, Q)$-Brownian motion, and, therefore, Q must be an ELMM, i.e., $\mathscr{H} \neq \emptyset$.

To prove the statement in the other direction, suppose that $Q \in \mathscr{H}$, but $\nu \not\sim \mathscr{W}$, i.e., there is a set $U \in \mathscr{A}_T$ with $\mathscr{W}(U) > 0$ and $\nu(U) \equiv \mu(\xi_{\boldsymbol{\cdot}} \in U) \equiv \mu(\phi^{-1}(U)) = 0$. Since $Q \in \mathscr{H}$, (ξ_t) must be a $(\mathscr{G}_t, Q)$-Brownian motion on $\mathbb{X}$. In particular, given any $A \in \mathscr{A}_T$, one must have $Q(\phi^{-1}(A)) = \mathscr{W}(A)$. However, since $Q \sim \mu$, it follows that $\mu(\phi^{-1}(U)) = 0$ is the same as $0 = Q(\phi^{-1}(U)) = \mathscr{W}(U)$, which contradicts to the choice of U, and shows that $\nu \not\sim \mathscr{W}$ is not possible when $\mathscr{H} \neq \emptyset$. □

3. THE TWO FUNDAMENTAL THEOREMS OF ASSET PRICING

Throughout this section we retain the notation, definitions, and assumptions introduced in Section 2. The main result in this paper is the following:

PROPOSITION 3.1. *If $\mathscr{H} = \emptyset$, or, equivalently, if $\nu \not\sim \mathscr{W}$, then guaranteed arbitrage exists; if $\mathscr{H} \neq \emptyset$, or, equivalently, if $\nu \sim \mathscr{W}$, then arbitrage is not possible.*

From the last result we immediately obtain the following:

COROLLARY 3.2. *The existence of arbitrage implies the existence of guaranteed arbitrage.*

Proof of Proposition 3.1. The second statement in the proposition is well known and essentially trivial (the fact that in our setting the ELMM lives on a special space does not alter the usual argument). Nevertheless, for the sake of completeness, we prove both statements, and stress that the entire proof is confined to the realm of stochastic analysis. We prove the first statement in the proposition first.

Suppose that $\nu \not\sim \mathscr{W}$. Then there is a set $U \in \mathscr{A}_T$ with the properties: $\mathscr{W}(U) > 0$ and $\nu(U) \equiv \mu(\xi_{\boldsymbol{\cdot}} \in U) = 0$. Let $U \in \mathscr{A}_T$ be one such set and let $F : \mathbb{W} \mapsto \mathbb{R}$ be defined as

$$\mathbb{W} \ni w \rightsquigarrow F(w) := 1_{\mathbb{W}\setminus U}(w) - \frac{1 - \mathscr{W}(U)}{\mathscr{W}(U)} 1_U(w) \in \mathbb{R}\,.$$

With this choice we have $\mathtt{E}_{\mathscr{W}}[F] = 0$ and

$$-\frac{1 - \mathscr{W}(U)}{\mathscr{W}(U)} \leq \mathtt{E}_{\mathscr{W}}[F \mid \mathscr{A}_t] \leq 1, \quad \mathscr{W}\text{-a.e. in } \mathbb{W}, \quad \text{for every fixed } t \in [0, T]\,.$$

As is well known, the Wiener measure has the predictable representation property (PRP). Consequently, one can find a $(\mathscr{A}_t)$-predictable stochastic process $(h_t \in \mathbb{R})$, which is defined on $\mathbb{W}$, has the property

$$\int_0^T h_s(w)^2 \,\mathrm{d}s < \infty, \quad \text{for} \quad \mathscr{W}\text{-a.e. } w \in \mathbb{W},$$

and is such that the Itô-integral (taken with respect to the coordinate process on $\mathbb{W}$ and the Wiener measure $\mathscr{W}$), namely,

$$\Lambda_t(w) := \int_0^t h_s(w)\,\mathrm{d}w_s\,, \quad t \in [0, T]\,,$$

represents a continuous modification of the regular $(\mathscr{A}_t, \mathscr{W})$-martingale $(\mathrm{E}_{\mathscr{W}}[F \mid \mathscr{A}_t])_{t\in[0,T]}$. It is also clear that this modification can be chosen in such a way that

$$\inf\nolimits_{t\in[0,T]} \Lambda_t(w) \geqslant -\frac{1 - \mathscr{W}(U)}{\mathscr{W}(U)}, \quad \text{for} \quad \mathscr{W}\text{-a.e. } w \in \mathbb{W}\,. \tag{3.1}$$

Now consider the normalized excess-returns process (ξ_t), which is defined on $\mathbb{X}$ and has sample paths $\xi_{\boldsymbol{\cdot}}$ that are distributed in the space $\mathbb{W}$ with probability law ν (relative to the measure μ on $\mathbb{X}$). Since—see A.1—$\nu \prec \mathscr{W}$, we can claim that $(\Lambda_t(\xi_{\boldsymbol{\cdot}}))$ is a well-defined continuous stochastic process on $\mathbb{X}$, which, in fact, can be identified with a continuous version of the integral

$$\int_0^t h_s(\xi_{\boldsymbol{\cdot}})\,\mathrm{d}\xi_s\,, \quad t \in [0, T]\,,$$

understood as a stochastic integral on $\mathbb{X}$, taken with respect to the semimartingale (ξ_t) and defined under the measure μ. This identification is a consequence of the fact that convergence in probability relative to the Wiener measure $\mathscr{W}$ entails convergence in probability relative to the measure $\nu \prec \mathscr{W}$, and the fact that $(\xi_t(X_{\boldsymbol{\cdot}}))$ is a semimartingale in $(\Omega, (\mathscr{F}_t)_{t\in[0,T]}, P)$. For the trading strategy $\pi_t := {\alpha_t}^{-1} h_t(\xi_{\boldsymbol{\cdot}})$ we can write:

$$\int_0^t \pi_s\,\mathrm{d}\tilde{\chi}_s = \int_0^t h_s(\xi_{\boldsymbol{\cdot}})\,\mathrm{d}\xi_s = \Lambda_t(\xi_{\boldsymbol{\cdot}})\,,$$

the identities being understood to hold for all $t \in [0, T]$, everywhere outside some μ-negligible set in $\mathbb{X}$. Because of the estimate in (3.1) and the relation $\nu \prec \mathscr{W}$, the trading strategy (π_t) is obviously tame. At the same time, (π_t) is easily seen to represent a guaranteed arbitrage, since $\mu(\xi_{\boldsymbol{\cdot}} \in U) = \nu(U) = 0$, and therefore

$$\int_0^T \pi_s\,\mathrm{d}\tilde{\chi}_s = \Lambda_T(\xi_{\boldsymbol{\cdot}}) = F(\xi_{\boldsymbol{\cdot}}) = 1, \quad \mu\text{-a.e. in } \mathbb{X}\,.$$

Finally, we show that arbitrage is not possible if an ELMM $Q \in \mathscr{H}$ exists. We merely adapt to our setting the usual argument.

Suppose that $\mathscr{H} \neq \emptyset$ and that, at the same time, the trading strategy (π_t) represents arbitrage. Let $Q \in \mathscr{H}$ be arbitrarily chosen and let

$$A_+ := \left\{x \in \mathbb{X} : \int_0^T \pi_s(x)\,\mathrm{d}\tilde{\chi}_s(x) \geqslant 0\right\} \text{ and } A_{++} := \left\{x \in \mathbb{X} : \int_0^T \pi_s(x)\,\mathrm{d}\tilde{\chi}_s(x) > 0\right\}.$$

Since (π_t) represents arbitrage, we must have $\mu(A_+) = 1$ and $\mu(A_{++}) > 0$, and these two relations are the same as $Q(A_+) = 1$ and $Q(A_{++}) > 0$. Furthermore, since the trading strategy (π_t) is tame, the sample paths $t \rightsquigarrow \int_0^t \pi_s\,\mathrm{d}\tilde{\chi}_s$ must be bounded from below by some universal constant, μ-a.e. in $\mathbb{X}$, or, which is the same, Q-a.e. in $\mathbb{X}$. Consequently, the process $\int_0^t \pi_s \, \mathrm{d}\tilde{\chi}_s$, $t \in [0, T]$, must be a $(\mathscr{G}_t, Q)$-supermartingale starting from 0, and this contradicts to $Q(A_+) = 1$ and $Q(A_{++}) > 0$. □

Finally, we turn to the notion of completeness of the market (M, Θ, X) in the arbitrage-free case $\mathscr{H} \neq \emptyset \iff \nu \sim \mathscr{W}$. The discussion of this topic will be very brief. Our main objective is to show that the completeness of the market can be linked to the PRP, too.

We would like to know whether the family of payoffs

$$\left\{ \int_0^T \pi_s \, \mathrm{d}\tilde{\chi}_s : \pi \equiv \text{ tame trading strategy} \right\}$$

is dense in the space $L^0(\mathbb{X}, \mathscr{G}_T, \mu; \mathbb{R})$, assumed to be equipped with the topology of convergence in probability μ. Since an equivalent change of the probability law does not alter the convergence in probability, in the study of any property that can be described only in terms of convergence in probability μ we are free to replace μ with *any* ELMM $Q \in \mathscr{H}$. Clearly, the payoff associated with *any* trading strategy can be expressed as the limit in probability of payoffs associated with *tame* trading strategies. Furthermore, any element of $L^0(\mathbb{X}, \mathscr{G}_T, \mu; \mathbb{R})$ can be expressed as a limit in probability of bounded elements of $L^0(\mathbb{X}, \mathscr{G}_T, \mu; \mathbb{R})$. Thus, what we want to know is whether the family of all random variables of the form

$$\int_0^T \pi_s \, \mathrm{d}\tilde{\chi}_s,$$

for all possible choices of the $(\mathscr{G}_t)$-predictable process (π_t), with

$$\mathrm{E}_Q\left[\int_0^T {\pi_s}^2 {\alpha_s}^2 \, \mathrm{d}s\right] < \infty,$$

is dense in $L^2(\mathbb{X}, \mathscr{G}_T, Q; \mathbb{R})$, for every ELMM $Q \in \mathscr{H}$. Due to lemma (4.2), theorem (4.7), and the comment following definition (4.8) in Revuz and Yor (1999), ch. V, this is the same as wanting to know whether, given *any* ELMM $Q \in \mathscr{H}$, one can construct a nontrivial $(\mathscr{G}_t, Q)$-local martingale (L_t) on $\mathbb{X}$, with $\langle \tilde{\chi}, L \rangle_t \equiv 0$. It is also the same as wanting to know whether the excess-returns process $(\tilde{\chi}_t)$—recall that $(\tilde{\chi}_t)$ represents the first component of the canonical coordinate process on $\mathbb{X}$—has the $(\mathscr{G}_t, Q)$-PRP on $\mathbb{X}$, *for every* $Q \in \mathscr{H}$. Motivated by the above remarks, we adopt the following:

DEFINITION 3.3. If the financial market (M, Θ, X) is free of arbitrage, then it is said to be "complete," if the excess-returns process $(\tilde{\chi}_t)$ has the $(\mathscr{G}_t, Q)$-PRP, for every ELMM $Q \in \mathscr{H}$.

PROPOSITION 3.4. *If the market* (M, Θ, X) *is free of arbitrage, then it is complete if and only if the family* $\mathscr{H}$ *of all ELMM is a singleton.*

Proof. Since $\mathscr{G}_0$ is μ-a.e. trivial, by theorem (4.7) and the comment following definition (4.8) in Revuz and Yor (1999), ch. V, the market is complete if and only if every ELMM Q is extremal in $\mathscr{H}$, i.e., if and only if $\mathscr{H}$ is a singleton. □

It is not very difficult to characterize the completeness of the market as a property of the distribution laws of the market process (X_t) and the normalized excess-returns process $(\xi_t(\tilde{X}_\cdot))$—respectively, μ and ν. Plainly, if the market is arbitrage-free, then it is complete if and only if the normalized excess-returns process is *essentially invertible*, where "essentially invertible" is made precise in the following statement:

PROPOSITION 3.5. *If the market is free of arbitrage, then it happens to be complete if and only if the distribution law, ν, of the normalized excess-returns process (ξ_t), uniquely determines the distribution law, μ, of the market process (X_t), within the class of probability measures on $\mathbb{X}$ that are equivalent to the law of (X_t).*

Proof. We adopt the notation introduced in the proof of Proposition 2.3, and recall that the Borel mapping $\phi : \mathbb{X} \mapsto \mathbb{W}$ is an arbitrary version of the random element $\xi_\cdot$. Suppose first that the market is free of arbitrage and complete. We must show that if μ' is another Borel measure on $\mathbb{X}$, such that $\mu' \sim \mu$ and $\mu \circ \phi^{-1} = \mu' \circ \phi^{-1} = \nu$—i.e., μ' is equivalent to μ and ϕ has the same distribution, ν, under both μ and μ'—then one must have $\mu = \mu'$. To see why this claim can be made, consider the transition measure $\Pi'(w, A)$, which is obtained from ϕ and μ' in the same way in which the transition measure $\Pi(w, A)$ was obtained from ϕ and μ in the proof of Proposition 2.3. Just as before, we set

$$Q(A) := \mathrm{E}_{\mathscr{W}}[\Pi(w, A)] \quad \text{and} \quad Q'(A) := \mathrm{E}_{\mathscr{W}}\left[\Pi'(w, A)\right], \quad A \in \mathscr{G}_T,$$

and, by repeating the argument of the proof of Proposition 2.3, since $\nu \sim \mathscr{W}$ (as the market is assumed arbitrage-free), we conclude that Q and Q' are both ELMM. Since the market is also complete, we must have $Q = Q'$, and this means that for every $A \in \mathscr{G}_T$ one has

$$\Pi(w, A) = \Pi'(w, A), \text{ for } \quad \mathscr{W}\text{-a.e. } w \in \mathbb{W}.$$

Since the property "$\mathscr{W}$-a.e." is the same as "ν-a.e.," and since

$$\mu(A) = \mathrm{E}_\nu[\Pi(w, A)] \quad \text{and} \quad \mu'(A) = \mathrm{E}_\nu[\Pi'(w, A)], \text{ for any } A \in \mathscr{G}_T,$$

we conclude that $\mu = \mu'$.

Now we must prove the statement in the other direction: if ν uniquely determines μ within the equivalence class of μ, then there could be only one ELMM—recall that $\mathscr{H} \neq \emptyset \iff \nu \sim \mathscr{W}$. Let $Q \in \mathscr{H}$ and let $q(w, A)$ be the transition measure constructed from ϕ and Q in the same way in which $\Pi(w, A)$ was constructed from ϕ and μ; in particular, $Q(A) = \mathrm{E}_{\mathscr{W}}[q(w, A)]$, $A \in \mathscr{G}_T$. Given any $A \in \mathscr{G}_T$, let

$$A^*_Q := \{w \in \mathbb{W} : q(w, A) > 0\},$$

and notice that, since $Q \sim \mu$ and $\nu \sim \mathscr{W}$, one must have $\mu(A) = 0 \iff Q(A) = 0 \iff \mathscr{W}(A^*_Q) = 0 \iff \nu(A^*_Q) = 0$. Setting $\mu'(A) := \mathrm{E}_\nu[q(w, A)]$, $A \in \mathscr{G}_T$, the last observation entails $\mu \sim \mu'$. Furthermore,

$$\mu'[\phi \in B] = \mathrm{E}_\nu[1_B(w)] = \nu(B) = \mu[\phi \in B], \qquad \text{for any} \qquad B \in \mathscr{A}_T,$$

with the implication that ϕ has the same distribution, ν, under both μ and μ'. Therefore $\mu = \mu'$, which is the same as the claim that, given any $A \in \mathscr{G}_T$, one has

$$q(w, A) = \Pi(w, A), \quad \text{for} \quad \nu\text{-a.e. } w \in \mathbb{W}.$$

Since the property "ν-a.e." is the same as "$\mathscr{W}$-a.e.," the last relation implies that

$$Q(A) = \mathrm{E}_{\mathscr{W}}[\Pi(w,\ A)], \quad A \in \mathscr{G}_T\,.$$

As the expectation in the right-hand side above does not depend on Q, it follows that Q must be unique. The market is therefore complete. □

4. AN EXAMPLE: DETECTION OF ARBITRAGE

Within the class of Itô-type models of financial markets, the study of arbitrage in the special class of Itô-type *diffusion* models is somewhat easier. For example, theorem 7.7 in Liptser and Shiryaev (1974) shows that no arbitrage would be possible if the normalized excess-returns follow some diffusion process on the space $\big(\mathbb{X}, (\mathscr{G}_t)_{t\in[0,T]}, \mu\big)$, provided that the associated drift obeys some minor integrability conditions; in fact, as the first integrability condition in the aforementioned result in Liptser and Shiryaev (1974) happens to be a consequence of the nonexplosion of the economy, the second integrability condition is *necessary and sufficient* for the absence of arbitrage. Unfortunately, necessary and sufficient conditions of this type are difficult to obtain for general (i.e., nondiffusion type) Itô-models. The fundamental reason for this is that sufficient conditions for equivalence are readily available, whereas, generally, necessary conditions are not—except in the diffusion case. Typically, in nondiffusion-type Itô-models, proving that the market allows arbitrage is much harder than proving that the market is free of arbitrage. In what follows we shall illustrate with a concrete example how the tools developed in Section 3 may be helpful in this regard.

The market in this example has a finite time horizon, $T > 0$, and is comprised of a risk-free bond, which pays constant interest $r > 0$, and a risky security ($S_t \in \mathbb{R}_{++}$), with price dynamics governed by the relation

$$\frac{\mathrm{d}S_t}{S_t} = \sigma\,\mathrm{d}W_t + \sigma^2\frac{V_t}{S_t}\,\mathrm{d}t + \frac{2\phi}{V_t}\,\mathrm{d}t, \quad 0 \leqslant t \leqslant T,$$

in which $\sigma > 0$ and $\phi > 0$ are given constants, (W_t) is some one-dimensional Brownian motion, and the process ($V_t \in \mathbb{R}_{++}$) is defined as the solution to the equation (driven by the same Brownian motion (W_t)):

$$\mathrm{d}V_t = V_t(\sigma\,\mathrm{d}W_t + \sigma^2\,\mathrm{d}t) + \phi\,\mathrm{d}t, \quad 0 \leqslant t \leqslant T,$$

with initial value $V_0 = S_0 > 0$ (S_0 is a given constant). The excess-returns process for this market is

$$\tilde{X}_t = \sigma\,W_t + \int_0^t \left(\sigma^2\frac{V_s}{S_s} + \frac{2\phi}{V_s} - r\right)\mathrm{d}s, \quad 0 \leqslant t \leqslant T,$$

and the information process is ($\bar{X}_t \equiv V_t$). The normalized excess-returns process is therefore given by

$$\xi_t = W_t + \int_0^t \left(\sigma\frac{V_s}{S_s} + \frac{2\phi}{\sigma\,V_s} - \frac{r}{\sigma}\right)\mathrm{d}s, \quad 0 \leqslant t \leqslant T\,.$$

It is easy to check that the process $U_t := S_t - V_t$ satisfies the equation

$$\mathrm{d}U_t = U_t\left(\sigma\,\mathrm{d}W_t + \frac{2\phi}{V_t}\,\mathrm{d}t\right) + \phi\,\mathrm{d}t,$$

and that one can therefore write

$$U_t = e^{\sigma W_t + \int_0^t \frac{2\phi}{V_s}\,\mathrm{d}s - \frac{1}{2}\sigma^2 t}\int_0^t \phi\, e^{\sigma W_s + \int_0^s \frac{2\phi}{V_u}\,\mathrm{d}u - \frac{1}{2}\sigma^2 s}\,\mathrm{d}s\,.$$

Similarly, one can write

$$V_t = e^{\sigma W_t + \frac{1}{2}\sigma^2 t}\left(V_0 + \phi\int_0^t e^{-\sigma W_s - \frac{1}{2}\sigma^2 s}\,\mathrm{d}s\right)$$

and

$$\mathrm{d}\log(V_t) = \frac{1}{V_t}\,\mathrm{d}V_t - \frac{1}{2}\sigma^2\,\mathrm{d}t \equiv \sigma\,\mathrm{d}W_t + \frac{1}{2}\sigma^2\,\mathrm{d}t + \frac{\phi}{V_t}\,\mathrm{d}t,$$

and then conclude that

$$W_t = \frac{1}{\sigma}\log\left(\frac{V_t}{V_0}\right) - \frac{1}{2}\sigma\,t - \int_0^t \frac{\phi}{\sigma V_s}\,\mathrm{d}s\,.$$

Consequently, all quantities ξ_t, U_t, and V_t can be treated as nonanticipative functionals of the sample path $[0, T] \ni t \rightsquigarrow W_t$, which are defined *everywhere* in the space $\mathscr{C}([0, T]; \mathbb{R})$. Furthermore, the above-mentioned relation shows that the sample path $[0, T] \ni t \rightsquigarrow W_t$ also can be treated as a nonanticipative functional of $[0, T] \ni t \rightsquigarrow V_t(W_\cdot)$, i.e., the mapping $V : \mathscr{C}([0, T]; \mathbb{R}) \mapsto \mathscr{C}([0, T]; \mathbb{R})$ is invertible. Thus, observing the sample path $t \rightsquigarrow V_t$ is tantamount to observing the sample path $t \rightsquigarrow W_t$ and, as a result, the filtration generated by the market process $(\tilde{X}_t, \bar{X}_t \equiv V_t)$ is nothing but the Brownian filtration associated with (W_t).

Next, define the process

$$Z_t = \frac{1}{\sigma}\log\left(\frac{V_t}{V_0}\right) - \frac{1}{2}\sigma\,t + \int_0^t \frac{\phi}{\sigma V_s}\,\mathrm{d}s \equiv W_t + 2\int_0^t \frac{\phi}{\sigma V_s}\,\mathrm{d}s, \quad 0 \leqslant t \leqslant T,$$

and observe that, in terms (Z_t), the dynamics of (V_t) can be expressed as

$$\mathrm{d}V_t = V_t(\sigma\,\mathrm{d}Z_t + \sigma^2\,\mathrm{d}t) - \phi\,\mathrm{d}t,$$

while the dynamics of (U_t) can be expressed as

$$\mathrm{d}U_t = \sigma U_t\,\mathrm{d}Z_t + \phi\,\mathrm{d}t\,.$$

As a result, one can write:

$$\xi_t = Z_t + \int_0^t \left(\frac{\sigma V_s}{U_s + V_s} - r\right)\mathrm{d}s, \quad 0 \leqslant t \leqslant T,$$

$$V_t = e^{\sigma Z_t + \frac{1}{2}\sigma^2 t}\left(V_0 - \phi\int_0^t e^{-\sigma Z_s - \frac{1}{2}\sigma^2 s}\,\mathrm{d}s\right), \quad 0 \leqslant t \leqslant T,$$

and

$$U_t = \phi\, e^{\sigma\,\mathrm{d}Z_t - \frac{1}{2}\sigma^2 t}\int_0^t e^{-\sigma\,\mathrm{d}Z_s + \frac{1}{2}\sigma^2 s}\,\mathrm{d}s, \quad 0 \leqslant t \leqslant T\,.$$

We can now treat the quantities ξ_t, U_t, and V_t as nonanticipative functionals of the sample path $[0, T] \ni t \rightsquigarrow Z_t$. Furthermore, there is an obvious one-to-one correspondence between the sample paths $[0, T] \ni t \rightsquigarrow Z_t$ and $[0, T] \ni t \rightsquigarrow V_t$ (due to the one-to-one correspondence between $[0, T] \ni t \rightsquigarrow W_t$ and $[0, T] \ni t \rightsquigarrow V_t$). Since $V_t > 0$ by definition, one must have

$$\int_0^t e^{-\sigma Z_s - \frac{1}{2}\sigma^2 s}\, \mathrm{d}s < \frac{V_0}{\phi}, \quad \text{for all } t \in [0, T],$$

and therefore

$$\int_0^T e^{-\sigma \xi_s - \frac{1}{2}\sigma^2 s}\, \mathrm{d}s = \int_0^T e^{-\sigma Z_s - \frac{1}{2}\sigma^2 s}\, e^{-\sigma \int_0^s \left(\frac{\sigma V_u}{U_u + V_u} - r\right) \mathrm{d}u} ds < \frac{V_0}{\phi}\, e^{\sigma(\sigma + r) T}.$$

Thus, the entire range

$$\{\xi_{\cdot}(w) : w \in \mathscr{C}([0, T]; \mathbb{R})\}$$

must be contained in the set

$$A_T = \left\{ w \in \mathscr{C}([0, T]; \mathbb{R}) : \int_0^T e^{-\sigma w_s - \frac{1}{2}\sigma^2 s}\, \mathrm{d}s < \frac{V_0}{\phi}\, e^{\sigma(\sigma + r) T} \right\}.$$

It is not very difficult to see that the Wiener volume of this last set is strictly less than 1. Not only does Proposition 3.1 show that this financial market allows arbitrage, but it also provides a method for constructing a *guaranteed arbitrage strategy* from the predictable representation of an appropriate linear combination of the indicator of the set A_T and the indicator of its complement.

APPENDIX

Proof of 1.3. Consider the set

$$A := \left\{ \omega \in \Omega : \int_0^T {\theta_s}^2\, \mathrm{d}s = \infty \right\},$$

and suppose that, for some $\varepsilon \in [0, 1]$, one can claim that $P(A) \geqslant \varepsilon$. Given any integer $k > 0$, let $\theta_t^{(k)} := (k \wedge \theta_t) \vee (-k)$, but suppose further that $\theta_t^{(k)}$ is defined also for $t > T$ by $\theta_t^{(k)} := 1$. Then set

$$Y_t := B_t + \int_0^t \theta_s\, \mathrm{d}s, \quad t \geqslant 0,$$

and notice that

$$\int_0^t \theta_s^{(k)}\, \mathrm{d}Y_s = \int_0^t \theta_s^{(k)}\, \mathrm{d}B_s + \int_0^t \left(\theta_s^{(k)}\right)^2 \mathrm{d}s + \int_0^t \underbrace{\theta_s^{(k)}\left(\theta_s - \theta_s^{(k)}\right)}_{\geqslant 0}\, \mathrm{d}s.$$

If $Q_t^{(k)} := \int_0^t (\theta_s^{(k)})^2\, \mathrm{d}s$, then, by the Dambis–Dubins–Schwartz (DDS) result, we can write

$$\int_0^t \theta_s^{(k)}\, \mathrm{d}B_s + \int_0^t \left(\theta_s^{(k)}\right)^2 \mathrm{d}s = Z^{(k)}_{Q_t^{(k)}} + Q_t^{(k)}, \quad t \geqslant 0,$$

where $(Z_t^{(k)})$ is the DDS Brownian motion associated with the local martingale $\int_0^t \theta_s^{(k)}\, \mathrm{d}B_s$, $t \geqslant 0$. Next, observe that for any finite $a \geqslant 0$ one must have

$$\inf_{t\geqslant 0}\big(Z_{a+t}^{(k)} + (a+t)\big) = Z_a^{(k)} + a - \sup_{t\geqslant 0}\big(-Z_{a+t}^{(k)} + Z_a^{(k)} - t\big),$$

and recall that (exercise (3.12) in Revuz and Yor 1999, ch. II) $\sup_{t\geqslant 0}(-Z_{a+t}^{(k)} + Z_a^{(k)} - t)$ is distributed with exponential law of parameter 2. Obviously, $\sup_{t\geqslant 0}(-Z_{a+t}^{(k)} + Z_a^{(k)} - t)$ is distributed independently from $Z_a^{(k)} + a$, which has a Gaussian law of mean a and variance a. As a result of this observation, for any constant $\Phi > 0$, the constant a can be chosen (independently from k, but, possibly, depending on ε), so that

$$P\big[\inf_{t\geqslant 0}\big(Z_{a+t}^{(k)} + (a+t)\big) \geqslant \Phi\big] \geqslant 1 - {}^{\varepsilon}/6\,.$$

Furthermore, the constant $0 > C > -\infty$ can be chosen (independently from both k and a, but, possibly, depending on ε) so that

$$P\big[\inf_{t\geqslant 0}(Z_t^{(k)} + t) \geqslant C\big] \geqslant 1 - {}^{\varepsilon}/6\,.$$

Now we can choose (depending on a) the integer k so that $P[Q_T^{(k)} \geqslant a] \geqslant {}^{2\varepsilon}/3$, in which case we have $P[E^{(k)}] \geqslant {}^{\varepsilon}/3$, where $E^{(k)} \subset \Omega$ is the set

$$E^{(k)} := \big\{\inf_{t\geqslant 0}\big(Z_{a+t}^{(k)} + (a+t)\big) \geqslant \Phi\big\} \cap \big\{\inf_{t\geqslant 0}\big(Z_t^{(k)} + t\big) \geqslant C\big\} \cap \big\{Q_T^{(k)} \geqslant a\big\}.$$

To complete the proof, it is enough to notice that in the above set one has

$$Z_{Q_T^{(k)}} + Q_T^{(k)} \geqslant \Phi \quad \text{and} \quad \inf_{t\geqslant 0}\big(Z_{Q_t^{(k)}} + Q_t^{(k)}\big) \geqslant C\,,$$

so that the property stated in the proposition must hold with $h_t \equiv \alpha_t{}^{-1}\theta_t^{(k)} 1_{[0,\tau_k]}(t)$, where

$$\tau_k := \inf\big\{t \geqslant 0 : Z_{Q_t^{(k)}}^{(k)} + Q_t^{(k)} < C\big\}$$

and, for this stopping time, we have $\tau_k = \infty$ on the set $E^{(k)}$. □

PROPOSITION A.1 (Restatement of theorem 7.4 in Liptser and Shiryaev (1974)). *Consider the normalized excess-returns process*

$$\eta_t := \xi_t(X_\cdot) = B_t + \langle B, \Theta\rangle_t, \quad t \in [0, T],$$

as a process on the filtered space $(\Omega, (\mathscr{F}_t)_{t\in[0,T]}, P)$, *where* $(B_t \in \mathbb{R})$ *is a* $(\mathscr{F}_t, P)$*-Brownian motion and* $(\Theta_t \in \mathbb{R})$ *is some* $(\mathscr{F}_t, P)$*-local martingale with* $\langle\Theta,\Theta\rangle_T < +\infty$, *P-a.e. in* Ω. *Then the distribution law* ν, *of the process* $(\eta_t) \equiv (\xi_t(X_\cdot))$ *in the space* $\mathbb{W}$, *is absolutely continuous with respect to the Wiener measure* $\mathscr{W}$ *on that space, i.e.,* $\nu \prec \mathscr{W}$.

Proof. We merely adapt to our setting the argument used in the proof from Liptser and Shiryaev (1974). Given any integer $k > 0$, define the $(\mathscr{F}_t)$-stopping time on Ω

$$\tau_k := \inf\{0 < t \leqslant T : \langle\Theta,\Theta\rangle_t > k\}\,,$$

with the understanding that $\tau_k = T$ if the above set is empty. Suppose that $\langle\Theta,\Theta\rangle_T < +\infty$ everywhere outside some P-negligible set $\mathscr{N} \in \mathscr{F}_T$ (i.e., $\mathscr{N} \in \mathscr{F}_0$) and, given any integer $k > 0$, define the process

$$Y_t^{(k)} = B_t + \langle B, \Theta\rangle_{t\wedge\tau_k} \equiv B_t + \langle B, \Theta^{\tau_k}\rangle_t, \quad t \in [0, T].$$

It is clear that in the set $\Omega \setminus \mathscr{N}$ one has $\lim_{k\nearrow\infty}\tau_k = T$; in fact, in that set one has $\tau_k = T$, for all sufficiently large integers k. Consequently, for any $\omega \in \Omega \setminus \mathscr{N}$, the sample paths $t \rightsquigarrow Y_t^{(k)}(\omega)$ coincide with the sample path $t \rightsquigarrow \eta_t(\omega)$, for all sufficiently large integers k (how large may depend on $\omega \in \Omega \setminus \mathscr{N}$). As a result, for every Borel set $A \in \mathscr{A}_T$, one has

$$\lim_{k\nearrow\infty} 1_A(Y_\bullet^{(k)}(\omega)) = 1_A(\eta_\bullet(\omega)), \quad \text{for every } \omega \in \Omega \setminus \mathscr{N}.$$

Given any integer $k > 0$, let $Q^{(k)}$ denote the probability measure on Ω defined by

$$\frac{\mathrm{d}Q^{(k)}}{\mathrm{d}P} = \exp\left(-\Theta_{\tau_k} - \frac{1}{2}\langle\Theta, \Theta\rangle_{\tau_k}\right).$$

By Cameron–Martin–Girsanov's theorem, the process $(Y_t^{(k)})$ is a $(\mathscr{F}_t, Q^{(k)})$-Brownian motion for every k. Furthermore, the measures $Q^{(k)}$, $k \geqslant 1$, and P all share the same family of null-sets. In particular, given any $A \in \mathscr{A}_T$, $\mathscr{W}(A) = 0$ implies $\mathscr{W}(A) = Q^{(k)}[Y_\bullet^{(k)} \in A] = 0 \iff P[Y_\bullet^{(k)} \in A] = 0$. By the dominated convergence theorem,

$$P[\eta_\bullet \in A] \equiv \mathrm{E}_P[1_A(\eta_\bullet)] = \lim_{k\nearrow\infty} \mathrm{E}_P[1_A(Y_\bullet^{(k)})] \equiv \lim_{k\nearrow\infty} P[Y_\bullet^{(k)} \in A],$$

for every Borel set $A \in \mathscr{A}_T$. Thus, if $\mathscr{W}(A) = 0$, one must have $\nu(A) \equiv P[\eta_\bullet \in A] = 0$, since $\mathscr{W}(A) = 0$ implies that all terms under the limit in the right-hand side above must vanish. □

REFERENCES

DELBAEN, F., and W. SCHACHERMAYER (1994): A General Version of the Fundamental Theorem of Asset Pricing, *Math. Ann.* 300, 463–520.

DELBAEN, F., and W. SCHACHERMAYER (1995): The Existence of Absolutely Continuous Local Martingale Measures, *Ann. Appl. Probab.* 5(4), 926–945.

DELBAEN, F., and W. SCHACHERMAYER (1998): The Fundamental Theorem of Asset Pricing for Unbounded Stochastic Processes, *Math. Ann.* 312, 215–250.

DELBAEN, F., and W. SCHACHERMAYER (2010): *The Mathematics of Arbitrage*, Berlin: Springer-Verlag.

DYBVIG, P., and S. ROSS (1987): Arbitrage, Vol. 1 of *The New Palgrave: A Dictionary of Economics*, London: Macmillan, pp. 100–106.

HARRISON, J. M., and D. M. KREPS (1979): Martingales and Arbitrage in Multiperiod Securities, *J. Econ. Theory* 20(3), 381–408.

HARRISON, J. M., and S. R. PLISKA (1981): Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stoch. Proc. Their Appl.* 11(3), 215–260.

KABANOV, Y. M., and D. O. KRAMKOV (1994): No-Arbitrage and Equivalent Martingale Measures: An Elementary Proof of the Harrison-Pliska Theorem, *Theory Probab. Appl.* 39(3), 523–527.

KABANOV, Y. M., and C. STRICKER (2001): A Teacher's Note on No Arbitrage Criteria, in *Séminaire de Probabilités XXXV*, Vol. 1755 of *Lecture Notes of Mathematics*, Berlin: Springer-Verlag, pp. 149–152.

LEVENTAL, S., and A. V. SKOROHOD (1995): A Necessary and Sufficient Condition for Absence of Arbitrage with Tame Portfolios, *Ann. Appl. Probab.* 5(4), 906–925.

LIPTSER, R. S., and A. N. SHIRYAEV (1974): *Statistika Slučajnih Procesov* [Russian], Moscow: Nauka; English Translation: *Statistics of Random Processes*. Springer-Verlag.

LYASOFF, A. (2007): A Crash Course in Stochastic Integration, *Essays in Mathematics* 1:4, Exegetist.com.

PARTHASARATHY, K. R. (1977): *Introduction to Probability and Measure*, London: Macmillan.

REVUZ, D., and M. YOR (1999): *Continuous Martingales and Brownian Motion*, 3rd ed., Springer-Verlag.

MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 505–532

BOUNDARY EVOLUTION EQUATIONS FOR AMERICAN OPTIONS

DANIEL MITCHELL

McCombs School of Business, University of Texas at Austin

JONATHAN GOODMAN

Courant Institute of Mathematical Sciences, New York University

KUMAR MUTHURAMAN

McCombs School of Business, University of Texas at Austin

We consider the problem of finding optimal exercise policies for American options, both under constant and stochastic volatility settings. Rather than work with the usual equations that characterize the price exclusively, we derive and use boundary evolution equations that characterize the evolution of the optimal exercise boundary. Using these boundary evolution equations we show how one can construct very efficient computational methods for pricing American options that avoid common sources of error. First, we detail a methodology for standard static grids and then describe an improvement that defines a grid that evolves dynamically while solving the problem. When integral representations are available, as in the Black–Scholes setting, we also describe a modified integral method that leverages on the representation to solve the boundary evolution equations. Finally we compare runtime and accuracy to other popular numerical methods. The ideas and methodology presented herein can easily be extended to other optimal stopping problems.

KEY WORDS: optimal stopping, American options, stochastic volatility, early exercise boundary, free-boundary problem, dynamic grid.

1. INTRODUCTION

American style options provide the holder the right to trade an underlying asset for a specified strike price before a specified expiry time. Pricing and finding the optimal exercise policy, which is known to be a surface that partitions the domain into exercise and hold regions, are interrelated and are solved for by transforming them to differential equation problems. The resulting differential equation, along with boundary conditions, formulate a free-boundary problem and characterize the price of the option. An accurate computation of the solution to the free-boundary problem relies on an accurate representation of the boundary and an accurate treatment of its dynamics. Rather than work with the equation that characterizes the price evolution of an American security exclusively, one could potentially derive and use the equations that characterize the evolution of the free-boundary for computational purposes. This however has not been seen as a valuable

Manuscript received March 2011; final revision received February 2012.

Address correspondence to Daniel Mitchell, McCombs School of Business, University of Texas at Austin, Austin, TX 78712, USA; e-mail: daniel.mitchell@phd.mccombs.utexas.edu.

DOI: 10.1111/mafi.12002

method because the boundary evolution equation also depends implicitly on the price, which seemed to be challenging to handle efficiently.

In this paper we consider American options, in the Black–Scholes setting and in a stochastic volatility setting, and derive boundary evolution equations. We show how one can construct computational procedures that efficiently utilize these evolution equations to compute both the price and the optimal exercise policy of American options. The evolution equations tell us exactly how fast the exercise boundary should move in time. This speed is dependent on both the current level of the boundary and a mixed derivative of the price function at the boundary, resulting in a system of differential equations. By solving these equations simultaneously we can track both the optimal exercise policy and the price function.

A challenge in constructing a boundary evolution equation–based computational procedure, apart from that posed by the implicit dependence of the equation on the price, is in taming the errors that arise from having to choose amongst points on a predefined grid to represent the boundary. Hence, we first construct a computational procedure that works on the standard rectangular cartesian grid by allowing a boundary to float between grid points. Although the performance of this first step is very encouraging, one could potentially eliminate any error due to the grid and boundary mismatch by allowing the grid to adapt to the boundary rather than predefine it. To this extent we next construct an improved methodology that dynamically builds a nonlinear grid while solving the boundary evolution equations. Such a dynamic evolution while being a relatively complicated implementation, performs significantly better and becomes essential under stochastic volatility. For cases where an integral representation of the option price is available, as is the case for the Black–Scholes model, we could potentially use the representation for further efficiency in solving the boundary evolution equation. We demonstrate how this can be done too for the Black–Scholes case. We also provide numerical evidence that the methods constructed in this paper are faster and more accurate than other relevant numerical methods. More work could potentially be done to extend this to existing integral representations of American options with stochastic volatility.

The primary objective of the paper is to show that it is possible to construct efficient numerical methods that take advantage of boundary evolution equations to avoid common sources of error. The expressions for the dynamics of the boundary in the Black–Scholes setting was found in van Moerbeke (1975) then rediscovered, independently, in Goodman and Ostrov (2002), and they were extended to some multifactor models in Hayes (2006). The corresponding equations for the stochastic volatility case have been derived in this paper. American option pricing is probably the most popular example amongst a larger class of very similar problems known as optimal stopping problems. Although this paper focuses exclusively on American options, the arguments for deriving the boundary evolution equations and the computational methods that solve these equations can readily be extended to other optimal stopping problems.

In Section 2, we consider the classical Black–Scholes setting of constant volatility and present three methods which leverage on the boundary equation. We then compare these methods to other numerical methods and find that all three methods constructed perform better. Section 3 considers the stochastic volatility case with a setting that is general enough to encompass several popular stochastic volatility models. For numerical comparisons, we only consider the most popular Heston model and describe the dynamic grid-based method in this context. We then compare the method to other existing numerical methods and find improved performance.

1.1. Background and Previous Literature

A put option is a contingent claim that gives the owner the right, but not the obligation, to sell a share of stock (or any other asset) at a prespecified price. Throughout the paper we restrict discussion to the put option only because almost the same arguments and equations will hold for the call option, wherein the owner has the right to buy a share of stock at a prespecified price. Put options come in two main flavors, "European," where the owner can only exercise this right at one prespecified time (expiration), and "American," where the owner can exercise this right anytime before expiration. When valuing an American put option the crucial step is to find the optimal early exercise boundary, which indicates the circumstances under which the option should be exercised before it expires. While a closed form solution for the value of a European option with constant volatility was found in the classical paper by Black and Scholes (1973) and for one particular stochastic volatility model in the paper by Heston (1993), there is no known closed form solution for the value of an American option with constant or stochastic volatility.

For constant volatility there are two main classes of numerical methods that approximate the price of American options. The first class computes the expected value of the American's payoff under the risk neutral measure. This class usually consists of Monte Carlo and binomial methods, and these methods only find the price of the option for one particular price and time to expiration and are typically unable to compute the early exercise boundary efficiently. The second class rephrases the expected value as the solution to a free-boundary partial differential equation (PDE), and methods in this class find the entire pricing function and the early exercise boundary. It can be difficult to compare methods in different classes because PDE methods give much more information than the first class.

The most well-known numerical method for solving the free-boundary problem was developed by Brennan and Schwartz (1977), but there have been several numerical methods developed since then. Muthuraman (2008) uses an iterative method to convert the free-boundary problem into a sequence of fixed boundary problems. Also Goodman and Ostrov (2002) find a differential equation that governs the early exercise boundary, which will be used heavily in this paper, and use it to derive a short-time asymptotic expansion of the boundary.

There are other methods that do not solve the free-boundary problem; rather they evaluate the risk neutral expected value of the option's payoff. Two common methods in practice that solve this problem are the binomial and trinomial tree methods. The binomial method was first seen in Cox, Ross, and Rubinstein (1979). Also, there has been much success in solving this problem using Monte Carlo simulation, most notably by Tilley (1993), Broadie and Glasserman (1997), and Longstaff and Schwartz (2001). Other methods that solve this problem partition the price as a European option's price plus an early exercise premium which results in an integral equation (Kim 1990; Jacka 1991; Carr, Jarrow, and Myneni 1992).

Recently there has been some work that exploits asymptotic analysis of the early exercise boundary to find approximate closed form solutions to the American option problem in the Black–Scholes setting. In Bunch and Johnson (2000) the authors find an implicit equation that can approximate the boundary at any time and then use numerical integration to find the price of the option. In Stamicar, Sevcovic, and Chadam (1999) the authors find an approximate explicit formula for the early exercise boundary. In Chen and Chadam (2007) the authors provide a detailed mathematical analysis of the early exercise boundary and provide an implicit ordinary differential equation (ODE) that governs the boundary. Evans, Kuske, and Keller (2002) provide results for American options on dividend paying stocks. A more comprehensive comparison of numerical methods can be found in AitSahlia and Carr (1997).

In the years since the seminal work of Black and Scholes there have been many empirical studies that suggest that simple Geometric Brownian Motion does not capture enough of the dynamics of a stock price to give an accurate price for derivative securities. As a result people have studied the case when the volatility of the stock follows a stochastic process. There have been several models that incorporate this but most work has focused on European options. As in the Black–Scholes setting, there is no known closed form solution for American options under any model.

Despite the vast research in American options with constant volatility there has been relatively less work exploring stochastic volatility. While some of the methods mentioned above can be extended to handle stochastic volatility, namely the PDE and Monte Carlo methods, there are also many methods that cannot handle stochastic volatility. Given the limitations of the above methods there has been some work looking for fast methods to price American options with stochastic volatility, including the multigrid method in Clarke and Parrott (1999) and a moving boundary method by Chockalingam and Muthuraman (2010). Ikonen and Toivanen (2007) use a componentwise splitting method to create three simple linear complementarity problems which they solve using the Brennan–Schwartz method. Also Wilmott (1998) describes how to use projected successive over relaxation (PSOR) to solve the free-boundary problem. In Detemple and Tian (2002) the authors present an integral representation for American options with stochastic volatility and interest rates that can be recursively solved to find the early exercise boundary. Broadie et al. (2000) use nonparametric techniques to investigate properties of the early exercise boundary under stochastic dividends and volatility. In Ikonen and Toivanen (2008) the authors present a more exhaustive review of other computational methods for American options with stochastic volatility.

2. CONSTANT VOLATILITY

In this section we generalize a boundary evolution equation, for the Black–Scholes setting, found in Goodman and Ostrov (2002), to a more general setting than nondividend paying stocks that includes assets such as futures, dividend paying stocks, and options on foreign currency. We then develop three numerical methods that leverage on the boundary evolution equation to obtain fast and accurate approximations of the price of an American option with constant volatility. For the rest of the paper we use the notation of Karatzas and Schreve (1998).

2.1. The Boundary Equation

We start with the classical Black–Scholes PDE for valuing an American put option, $p(x, \tau)$, where x is the price of the underlying asset and τ is the time until expiry. An American put option can be exercised at any time before it expires with payoff of $q - x$, where q is the strike price of the option. This suggests that we should partition the domain into two distinct regions separated by the early exercise boundary, $c(\tau)$. If at time τ, $x \leq c(\tau)$ the option should be exercised immediately with a payoff of $q - x$, and if $x > c(\tau)$ the option should be held. The optimal choice of $c(\tau)$ is decided by comparing the intrinsic value of the option to its tradable value; if it is worth more on the open market than its intrinsic value, then it should not be exercised. In the constant volatility case if $x > c(\tau)$ then $p(x, \tau)$ is governed by the classical Black–Scholes PDE,

$$\frac{\partial p}{\partial \tau} = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 p}{\partial x^2} + bx\frac{\partial p}{\partial x} - rp. \tag{2.1}$$

Here r is the risk-free interest rate, σ is the volatility of the underlying asset, and b is the instantaneous cost of carrying the underlying asset, as in Huang, Subrahmanyam, and Yu (1996). Using this notation for b allows us to price several financial instruments. For example, for nondividend paying stocks, $b = r$; for stocks with constant dividend yield δ, $b = r - \delta$; for futures, $b = 0$; and for options on foreign currency with foreign risk-free rate r_f, $b = r - r_f$.

We know that at $\tau = 0$ the option expires, thus it must be exercised or abandoned and therefore $c(0) = q$, if $b \geq 0$, otherwise $c(0) = \frac{r}{r-b}q$. The last thing we need to know about this option is the smooth pasting condition, which states that on the boundary p must be differentiable as shown in Merton (1992). With this information we can establish initial and boundary conditions for p, which are

$$p(x, 0) = \max(q - x, 0),$$

$$p(c(\tau), \tau) = q - c(\tau), \tag{2.2}$$

$$\frac{\partial p(c(\tau), \tau)}{\partial x} = -1, \tag{2.3}$$

and

$$\lim_{x\to\infty} p(x, \tau) = 0. \tag{2.4}$$

Equation (2.4) implies that

$$\lim_{x\to\infty} \frac{\partial p}{\partial x} = 0, \tag{2.5}$$

because p is convex and decreasing, as seen in Karatzas and Shreve (1998). It is more convenient, numerically, to use equation (2.5) as a boundary condition for large x, so we will not use equation (2.4) in numerical experiments.

Now that we have the boundary conditions we would like a differential equation that governs $c(\tau)$. We find this using higher-order derivatives that are continuous up to the early exercise boundary from the continuation region, but not across into the exercise region (see, for example, Lawrence and Salsa 2008 for a proof of this in several multiasset cases). We use these expressions to treat the price and boundary as a coupled system to be solved simultaneously.

THEOREM 2.1. *The differential equation that governs $c(\tau)$ is*

$$\frac{\partial c(\tau)}{\partial \tau} = -\frac{\partial^2 p(c(\tau), \tau)}{\partial x \partial \tau} \frac{\sigma^2 c^2(\tau)}{2qr - 2(r - b)c(\tau)}. \tag{2.6}$$

Proof. First differentiate the boundary conditions with respect to time, which will be in terms of the time derivative of p. This will also lead to time derivatives of the exercise boundary, $c(\tau)$, which is indeed differentiable by lemma 4.1 in Myneni (1992). Notice that the time derivative of p also satisfies equation (2.1). The first boundary condition, (2.2) becomes

$$\frac{\partial p}{\partial x}\frac{\partial c}{\partial \tau} + \frac{\partial p}{\partial \tau} = -\frac{\partial c}{\partial \tau}$$

which is simplified using equation (2.3) to

$$\frac{\partial p(c(\tau), \tau)}{\partial \tau} = 0. \tag{2.7}$$

Next take the time derivative of equation (2.3) and find

$$\frac{\partial^2 p}{\partial x^2}\frac{\partial c}{\partial \tau} + \frac{\partial^2 p}{\partial x \partial \tau} = 0. \tag{2.8}$$

Now take the limit as $x \to c(\tau)$ from the right and substitute equation (2.7) into equation (2.1) and get

$$0 = \frac{1}{2}\sigma^2 c^2(\tau)\frac{\partial^2 p(c(\tau), \tau)}{\partial x^2} + bc\frac{\partial p(c(\tau), \tau)}{\partial x} - rp(c(\tau), \tau). \tag{2.9}$$

Next substitute equations (2.2) and (2.3) into (2.9),

$$\frac{\partial^2 p(c(\tau), \tau)}{\partial x^2} = \frac{2qr - 2(r-b)c(\tau)}{\sigma^2 c^2(\tau)}. \tag{2.10}$$

Finally combine equations (2.8) and (2.10) and rearrange terms to find the desired function, (2.6). □

Even with this equation, finding the price of the American put is still a hard problem. We see in equation (2.6) that the boundary's evolution depends on both a mixed derivative of the price function and the current boundary level, which creates a system of nonlinear differential equations. The price of the put option depends on the boundary and the boundary depends on the price of the put. To solve these equations we must find a way to evolve them simultaneously.

Figure 2.1 shows the state space partitioned into the exercise region and the continuation region. The two regions are separated by the early exercise boundary. In the exercise region the price of the put is equal to its intrinsic value. In the continuation region the price of the put is governed by equation (2.1).

2.2. Numerical Method on a Static Grid

This section constructs a numerical method that uses equation (2.6) to compute the early exercise boundary and the price function of an American put option. The basic idea is to step forward in time to expiry discretely, evolving p and c at each step using finite difference approximations to equations (2.1) and (2.6). In this process several intricacies need to be addressed so we describe the algorithm with a three-step iterative procedure.

Step 1: Initialize *p* and *c* at a small time before expiration. To begin evolution using equations (2.1) and (2.6) we need an initial value of p and c. We know that at $\tau = 0$ the boundary is located at $c(0) = \min(q, \frac{r}{r-b}q)$, and at every value of x such that $x \geq q$ we have $p(x, 0) = 0$. In this numerical method we only consider the domain where $x \geq c(\tau)$

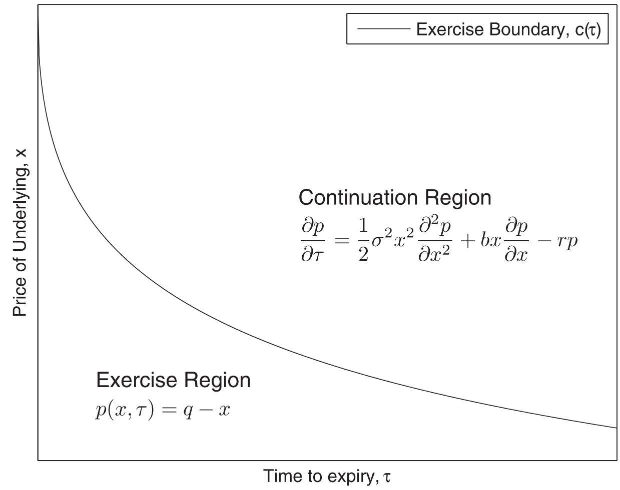


FIGURE 2.1. Partitioned state space.

because when $x > c(\tau)$ we know that p is governed by equation (2.1) and when $x = c(\tau)$, p must obey the boundary conditions so any value of $x < c(\tau)$ cannot be used. Using this information as an initial value we will see that all finite difference approximations to derivatives in the x variable will be zero if $b \geq 0$. This happens because we do not consider the domain such that $x < c(0)$, the only place where $p(x, 0) \neq 0$, so any place that we calculate a derivative in x will result in a linear combination of zeros, which is zero. For example, if $q = 100$ and $b > 0$ then $c(0) = q$. Now if we try to approximate the derivative of p at $x = 101$, using a finite difference approximation, we will find that $p(101 + \Delta x, 0) = p(101 - \Delta x, 0) = 0$, and $\frac{\partial p}{\partial x} \approx \frac{p(101+\Delta x,0)-p(101-\Delta x,0)}{2\Delta x} = \frac{0-0}{2\Delta x} = 0$.

This fact together with equation (2.1) tells us also that the numerical approximation of $\frac{\partial p}{\partial \tau}$ must also be zero. This means the price of the put cannot change in one step, and thus the location of the boundary cannot change in one step, if we start with the initial data at $\tau = 0$. If they do not move in the first step then they will not move in any subsequent step and the price of the put will stay at zero for all times to expiry, which is clearly incorrect. However, if $b < 0$ we do not have a problem.

To overcome this problem we approximate the put, at a short time before expiration, as a European option, as in Broadie and Detemple (1996), and with this we can take advantage of the closed form Black–Scholes equation for European options. On a discrete set of equally spaced grid points in x between zero and some $\hat{x}$, where $\hat{x}$ is the maximal value in the computational domain, we find the value of a European put, $f(x, \tau_o)$, a short time before expiration, τ_o. The choice of $\hat{x}$ is not entirely trivial here, we need to pick $\hat{x}$ so that equation (2.5) is approximately true for all times to expiry that we consider. Since f is a European option it must satisfy $f(x, \tau_o) = qe^{-r\tau_o}N(-d_2) - xe^{-(r-b)\tau_o}N(-d_1)$, where $d_1 = \frac{\log(x/q)+(b+\frac{1}{2}\sigma^2)\tau_o}{\sigma\sqrt{\tau_o}}$ and $d_2 = d_1 - \sigma\sqrt{\tau_o}$. Here N is the standard normal cumulative distribution function. We initialize p for a short time as $p(x, \tau_o) = \max(f(x, \tau_o), q - x)$.

To find the initial value for the boundary, $c(\tau_o)$, we use a binary search to find the place where $f(x, \tau_o)$ intersects the line $q - x$. Here we will almost certainly find that $c(\tau_o)$ is not located at one of the grid points chosen above, but this is not a problem; we will evolve p on the fixed grid and let c move between the grid points.

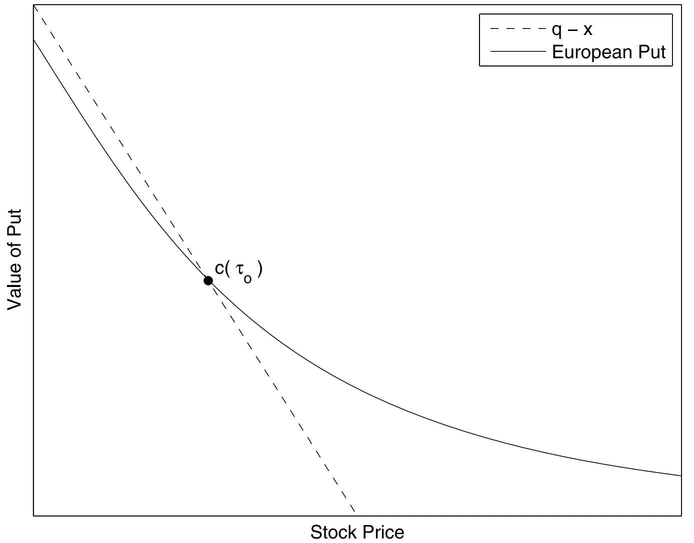


FIGURE 2.2. Initialization of $p(x, \tau_o)$ and $c(\tau_o)$.

In Figure 2.2, we illustrate how to initialize p and c. In the figure the dashed line represents the intrinsic value of the put and the solid line represents the value of a European option. We say that to the left of the intersection the American is equal to the dashed line, to the right of the intersection the American is equal to the solid line and the boundary is located at the intersection of the two lines. However we only work in the domain such that $x \geq c(\tau)$ so we only need the location of the intersection and the solid line to the right for initialization. Figure 2.2 exaggerates the initialization procedure for illustrative purposes. In the numerical experiments we run in Section 2.5 we find that the slope of the European option at the initial approximate boundary ranges between -0.999 and -0.975 when we initialize at half a trading day before expiration, $\tau_0 = \frac{1}{2}\frac{1}{252}$.

Step 2: Evolve *p* one step in time to expiry and approximate the mixed derivative. The next step is to evolve p one step backwards in time, holding $c(\tau)$ fixed. This, however, presents a problem because the values of p are not exactly uniform and we want to use a finite difference method. The grid points where we know p are uniformly spaced, but we also know the value of p at the boundary, which does not fit on this uniform spacing. To use a finite difference method we need to approximate all derivatives using the value of the price function at discrete grid points. For most of the grid points we can use standard central difference methods, however at the first grid point to the right of the boundary, call this point x_0, we cannot use these standard central difference formulae. To find the derivatives of p at x_0 we use Taylor series expansion to derive noncentral finite difference approximations involving x_0, $x_0 + h$ and $x_0 - h_2$. Here h_2 is the distance between x_0 and $c(\tau)$, and $p(x_0 - h_2) = q - c(\tau)$ because $x_0 - h_2 = c(\tau)$. One advantage of using this method to compute the derivatives of p at x_0 is that we can insert these equations directly into any discrete time stepping finite difference algorithm, like the Crank and Nicolson (1947) algorithm, which we use.

In the evolution of p we do not need to calculate any derivatives at $c(\tau)$ or $\hat{x}$ because we can use equations (2.2) and (2.5) as boundary conditions. Equation (2.2) means that in one time step the value of the put at $c(\tau)$ does not change. Equation (2.5) means that the value of the put at $\hat{x}$ is equal to the value of the put at the grid point just before $\hat{x}$. Both of these boundary conditions can easily be satisfied implicitly using the Crank–Nicolson algorithm.

After we evolve p we need to approximate the x derivative of the price function at the early exercise boundary so that we can use it to calculate $\frac{\partial^2 p(c(\tau),\tau)}{\partial x \partial \tau}$. To calculate this derivative we need to use the location of the boundary, $c(\tau)$, the value of the put at the boundary, $q - c(\tau)$, and a few grid points to the right of the boundary. In this calculation we cannot simply use standard one-sided finite difference formulae because the boundary is not located at a grid point. This means that the places where we know the value of the put to the right of the boundary are not equally spaced; if the distance between grid points is h then the space between the boundary and the first grid point to the right of the boundary must be less than h. To overcome this problem we fit a spline through the boundary and a few grid points to the right of the boundary. Using the coefficients of this spline we can analytically approximate the derivative of the price function at the boundary.

With this value for the x derivative we can approximate $\frac{\partial^2 p(c(\tau),\tau)}{\partial x \partial \tau}$ using a first-order finite difference method in time. If we say the value of the x derivative at the boundary before we evolved p is $p_x^{\text{old}} = -1$ and the value after we evolved p is p_x^{new} then $\frac{\partial^2 p(c(\tau),\tau)}{\partial x \partial \tau} \approx \frac{p_x^{\text{new}}+1}{\Delta\tau}$, where $\Delta\tau$ is the step size in time to expiry.

Step 3: Evolve c one step in time to expiry. Now that we have evolved p and calculated the mixed derivative at the boundary we need to evolve $c(\tau)$ one step in time to expiry to catch up with p. For this we hold $\frac{\partial^2 p(c(\tau),\tau)}{\partial x \partial \tau}$ fixed and use equation (2.6) and an explicit Runge–Kutta method to evolve c, we use the second-order Runge–Kutta method in numerical experiments; see Iserles (2008) for details on Runge–Kutta methods.

There is one last problem we face: what happens when the boundary crosses from one side of a grid point to the other? For example, if at time τ, $c(\tau)$ is located between the 99th and 100th grid points and at time $\tau + \Delta\tau$, $c(\tau)$ is located between the 98th and 99th grid points, what do we do? Here there are a few options as well. In Chen et al. (1997) the authors suggest using a spline to interpolate the value of p, however we find that simply leaving p at this point as its intrinsic value, $q - x$, does not lead to any significant reduction in accuracy, and thus using a spline is not worth the added complexity. We then repeat Steps 2 and 3 until we reach the desired time before expiration.

2.3. Numerical Method on a Dynamic Grid

In the previous section we allowed the optimal early exercise boundary to move between fixed grid points in the computational domain, however this can lead to some error when the boundary is very close to the next grid point and h_2 is very small when compared to h. To overcome this error we use numerical grid generation to force the grid points to conform to the boundary at every time step by using a change of variables. This forces the grid points we use, to approximate p, to move over time, and the space between the boundary and the closest grid point remains a constant. This has the advantage of allowing us to use standard high-order finite difference approximations when calculating the mixed derivative at the boundary and the standard difference methods at the first grid point greater than the boundary.

Numerical grid generation is used to transform complicated computational domains, through a change of variables, to much simpler domains that allow the use of standard

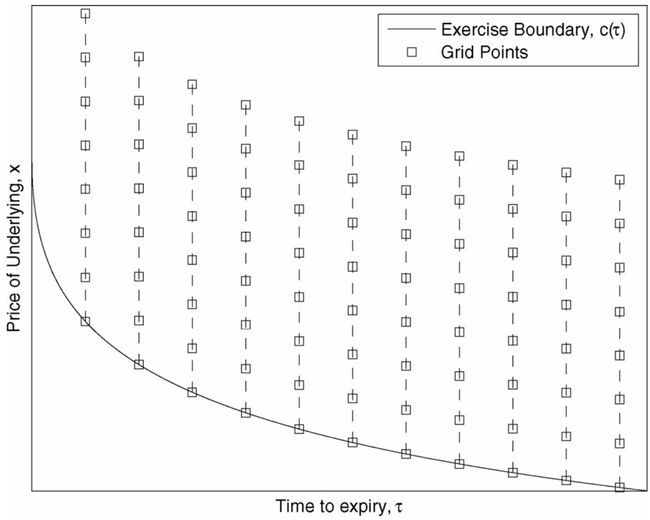


FIGURE 2.3. Computational grid in x at different times τ.

finite difference methods. In our case we wish to transform the domain $\{(x, \tau)\colon \forall\tau \geq 0, x \geq c(\tau)\}$ to $\mathbb{R}_+^2$. For details of numerical grid generation see Thompson, Warsi, and Mastin (1985). The front fixing methods in Nielsen, Skavhaug, and Tveito (2002) and Wu and Kwok (1997) also use a change of variables to eliminate the moving boundary, however this does not translate to a computational advantage because they do not use the boundary evolution equation considered here. The change of variable we use to transform our domain is

$$\omega = x - c(\tau), \tag{2.11}$$
$$g(\omega, \tau) = p(x, \tau).$$

Here ω can be interpreted as distance to the boundary. Given this change in variable we discretize ω uniformly from zero to some $\hat{\omega}$, which is equivalent to re-discretizing x at every time step uniformly from $c(\tau)$ to some $\hat{x}$, where $\hat{x}$ changes at each step to $c(\tau) + \hat{\omega}$. Figure 2.3 shows the computational grid in the (x, p) space for different values of τ. As τ increases, $c(\tau)$ decreases and the grid points align with the boundary for every value of τ. After transformation to the (ω, g) space the computational grid is a standard rectangular region.

Using the chain rule we find the PDE that governs the evolution of g and c to be

$$\frac{\partial g}{\partial \tau} = \frac{1}{2}\sigma^2(\omega + c(\tau))^2 \frac{\partial^2 g}{\partial \omega^2} + b(\omega + c(\tau))\frac{\partial g}{\partial \omega} - rg + \frac{\partial g}{\partial \omega}\frac{\partial c}{\partial \tau}, \tag{2.12}$$

$$\frac{\partial c(\tau)}{\partial \tau} = -\frac{\partial^2 g(0, \tau)}{\partial \omega \partial \tau}\frac{\sigma^2 c^2(\tau)}{2qr - 2(r - b)c(\tau)}. \tag{2.13}$$

The main difference between equations (2.12) and (2.1) is the addition of the final nonlinear term which comes from using the chain rule to differentiate g with respect to time. Known as the grid speed, this term allows us to find the value of p at the new grid points without the need for any sort of interpolation. Here, however, it is not as

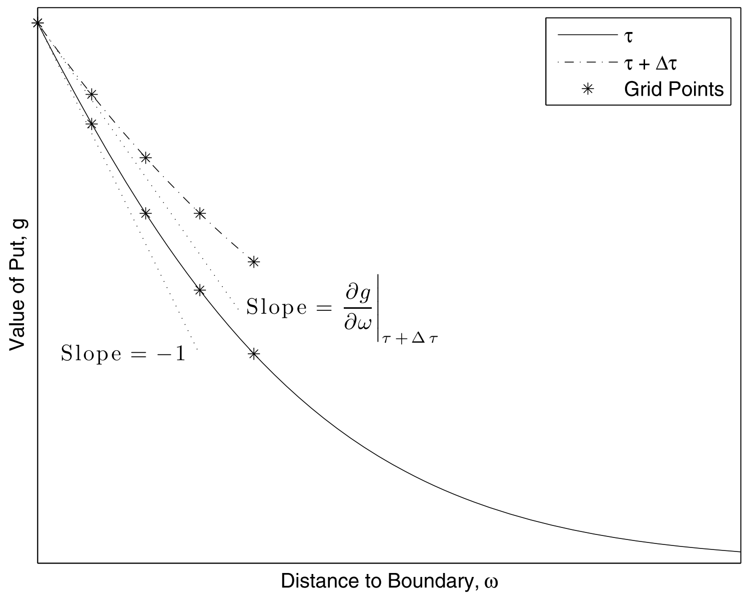


FIGURE 2.4. Difference in slopes, to calculate $\frac{\partial^2 g}{\partial\omega\partial\tau}$.

easy to estimate the mixed derivative at the boundary and we must come up with a new method of approximation. As the grid speed term in equation (2.12) also depends on the boundary evolution equation we cannot simply evolve g one step and use that to calculate the boundary evolution. The numerical method presented here can be described by a three-step iterative procedure as well.

Step 1: Initialization. We initialize p for a small time before expiry, τ_0, the same way we did in Section 2.2. However in this case we first find $c(\tau_0)$ and then initialize p uniformly between $c(\tau_0)$ and $\hat{x}$. Then we assign these values to ω and g according to the change of variables (2.11).

Step 2: Calculate $\frac{\partial^2 g(0,\tau)}{\partial\omega\partial\tau}$. Here, again, the calculation of the mixed derivative at the boundary can be difficult. To approximate this we simply evolve a few grid points greater than the boundary using equation (2.12) without the last term, the grid speed, and calculate the ω derivative after this evolution to use in a time finite difference method. Omitting the grid speed term has the effect of freezing the grid points for one small step in time and telling us how much the price of the put would change on those fixed grid points. For example if we use a 4-point finite difference approximation for the ω derivative and a second-order Runge–Kutta method for the time derivative, we only need to evolve six grid points larger than the boundary, so that we do not need to worry about right boundary conditions, which is not computationally expensive so we use this in numerical experiments.

Figure 2.4 shows how we calculate $\frac{\partial^2 g(0,\tau)}{\partial\omega\partial\tau}$ in the $(\omega,\ g)$ space. We see that all grid points are equally spaced and the grid points on the solid and dashed lines correspond to the same horizontal values because we dropped the grid speed term to get the dashed line. We also see that we only have the value of the put on the dashed line for a few grid points. Using the grid points on the dashed line and a standard one-sided finite difference equation we calculate the ω derivative. Using these two values with equation (2.3) we can approximate $\frac{\partial^2 g(0,\tau)}{\partial\omega\partial\tau}$.

Step 3: Evolve g and c simultaneously one step. Once we have calculated the mixed derivative we hold it constant while we evolve equations (2.12) and (2.13). Since we hold this constant we can use a coupled Runge–Kutta method to evolve g and c; in numerical experiments we use the second-order coupled Runge–Kutta method. Also for this method we use the same boundary condition for large ω as we did in Section 2.2. However to increase accuracy we add an extra grid point to the end of the computational domain every time $c(\tau) + \hat{\omega} < \hat{x}$. When this extra grid point is brought into the computational domain it is introduced according to equation (2.5). We repeat Steps 2 and 3 until we reach the desired time and then change the variables back to x and p.

We could potentially evolve the system implicitly but equations (2.12) and (2.13) are both nonlinear. This means we either need to linearize the equations or use a nonlinear solver to evolve the system, however we do not want to rely on the speed of any specific nonlinear solver to determine the computational time of the algorithm. When we consider stochastic volatility we will have nonlinear PDEs similar to these and we apply a linearization to the system. However in constant volatility the linearization is not beneficial on a fine mesh so we only use an explicit method.

In this method we cannot evolve g one step, calculate the mixed derivative, and then evolve c one step as we did in the previous section because the grid is moving. Considering the (x, p) space in the evolution of the price with the grid speed term, the grid points used to calculate the space derivative before the price evolution and the grid points used after the evolution are not the same. Therefore we cannot combine these values to calculate the mixed derivative. Also, the value of the grid speed term is partially determined by the mixed derivative. If we do not know the value of the mixed derivative, then we do not know the value of the grid speed term and we cannot evolve g through time.

Figure 2.5 shows the evolution of the value of the put and the grid points used to calculate its value over a step in time to expiry. The solid line represents the value of the put in the (x, p) space and the corresponding grid points in x before the price and boundary are evolved one step. The dashed line shows the value of the put and the corresponding grid points after the price and boundary are evolved in Step 3. We can see that the grid points on the solid and dashed curves do not coincide because they have moved over the course of a step in time to expiry. The first grid point on each line corresponds to the early exercise boundary at that time.

There are a couple of minor drawbacks to this algorithm. The first problem is that we will almost always have to use a spline to compute the price of the put at some x value after the algorithm is finished because we cannot pick the grid to include that value like we could in the method presented in Section 2.2. For example, if we wish to know the price of the option when the underlying stock costs \$100 the method in Section 2.2 lets us pick 100 to be a grid point since the grid is static. However if we choose the grid spacing so that 100 is a grid point when we initialize we will almost certainly find that 100 is not a grid point when the algorithm is finished because the grid has moved. Therefore, we must interpolate to find the value of the option when the underlying costs \$100. The next problem is that since we include the nonlinear term to the end of equation (2.12) the CFL condition forces us to use a smaller time step size than that required in Section 2.2. We will see however that despite these problems this method compares favorably in speed and accuracy to the static grid method.

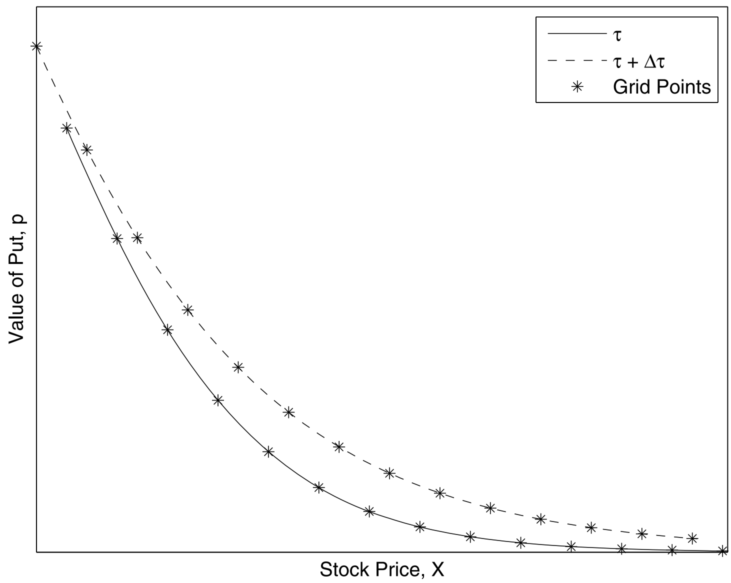


FIGURE 2.5. Illustrating different grid points used at two different times.

2.4. Modified Integral Method

In the previous sections we needed to evolve the boundary and the value function simultaneously because the boundary evolution equation requires a mixed derivative of the value function evaluated at the boundary. In this section, we present a numerical method that does not require the value function to be explicitly evolved with the boundary. This is achieved by using the integral representation of the American put option, as in Kim (1990). Although we do not directly extend this to American options with stochastic volatility, this method gives a good example of how to use boundary evolution equations to improve other numerical methods besides PDE methods.

The integral representation of the American put option states that the value of an American put is equal to the value of a European put, plus an early exercise premium. The early exercise premium is an integral of a function of the boundary. The value of an American put is

$$(2.14)\quad p(x,\tau) = f(x,\tau) + \int_0^{\tau} \left[rqe^{-r(\tau-u)}N(-d_2^*) - (r-b)xe^{-(r-b)(\tau-u)}N(-d_1^*) \right] \mathrm{d}u,$$

where

$$d_1^* = \frac{\log(x/c(u)) + \left(b + \frac{1}{2}\sigma^2\right)(\tau - u)}{\sigma\sqrt{\tau-u}} \quad \text{and} \quad d_2^* = d_1^* - \sigma\sqrt{\tau - u}.$$

Here $f(x,\tau)$ is the value of a European put, N is the standard normal cumulative distribution function, and we see that d_1^* and d_2^* are functions of the boundary.

Using this representation, if we know the value of the boundary between $\tau = 0$ and some time to expiry, τ_1, then we would like to express the value of the mixed derivative at τ_1 as some integral of the known boundary, which we can approximate using numerical integration. If this is possible we can then calculate the value of the boundary at

$\tau_1 + \Delta\tau$ using equation (2.6), and an ODE solver. This is exactly what we will do, however there are a few subtleties that arise that make this process complicated so we describe the process in four steps below.

Step 1: Initialization. To start this algorithm we must know the value of the boundary at a small time to expiry, as we did in the previous sections, because to calculate the mixed derivative using equation (2.14) we need something to integrate. To initialize this method we tried two different approaches. First we tried the asymptotic expansions for small τ, found in several papers mentioned in Section 1.1. We also tried the method we have used in previous sections: find the intersection of the European option with the intrinsic value of the put. It is somewhat surprising, but in numerical experiments we find that the method of finding the intersection is about five times more accurate than the asymptotic expansions that we tried, so we initialize with the binary search method. Once we know the boundary, we do not need to calculate the price function as we did in the previous sections because we will evaluate the mixed derivative as an integral of the boundary.

Step 2: Calculate $\frac{\partial^2 p}{\partial x \partial \tau}$. To calculate the mixed derivative we differentiate equation (2.14) first with respect to x and then τ. The derivative with respect to x is

$$\frac{\partial p}{\partial x} = \frac{\partial f}{\partial x} - \int_0^\tau \left[\frac{rq}{x\sigma\sqrt{\tau - u}} e^{-r(\tau-u)} N'(d_2^*) - \frac{r-b}{\sigma\sqrt{\tau-u}} e^{-(r-b)(\tau-u)} N'(d_1^*) \right. \\ \left. + (r-b)e^{-(r-b)(\tau-u)} N(-d_1^*) \right] \mathrm{d}u. \tag{2.15}$$

When we differentiate equation (2.15) with respect to τ, using the Liebniz rule, and evaluate at $x = c(\tau)$ we find that $\frac{\partial^2 p}{\partial x \partial \tau} = \frac{\partial^2 f}{\partial x \partial \tau} + \infty - \infty$.

This does not mean that the derivative does not exist, rather it means that there is no analytical expression for the derivative, because there is no analytical expression for the integral in equation (2.15). This happens because the integrand in equation (2.15) blows up when $u = \tau$, even though it is still integrable.

To overcome this problem we employ numerical integration to calculate $\frac{\partial p}{\partial x}$, using equation (2.15), evaluated at $(c(\tau - \Delta\tau), \tau)$ and assume that $\lim_{u \to \tau} d_{1,2}^* = 0$. Then we can approximate the mixed derivative as

$$\left. \frac{\partial^2 p}{\partial x \partial \tau} \right|_{(c(\tau),\tau)} \approx \left(\left. \frac{\partial p}{\partial x} \right|_{(c(\tau - \Delta\tau),\tau)} + 1 \right) \Big/ \Delta\tau.$$

Here the "plus one" comes from the assumption that $\frac{\partial p}{\partial x}|_{(c(\tau-\Delta\tau),\tau-\Delta\tau)} = -1$.

The approximation of the integral in equation (2.15) is also not entirely straightforward because the integrand blows up when $u = \tau$, so any standard numerical approximation will undervalue the integral. To fix this we split the integral into two parts, first the integral from 0 to $\tau - \Delta\tau$ and then the integral from $\tau - \Delta\tau$ to τ. The first part of the integral can easily be approximated using any numerical integration technique and the second part of the integral can be approximated in closed form if we assume that $d_{1,2}^* = 0$ on the interval $[\tau - \Delta\tau, \tau]$. In fact, $d_{1,2}^* = 0$ when $u = \tau$ only if we evaluate at $x = c(\tau)$. Since we are trying to approximate the derivative at a value close to $c(\tau)$ we use $d_{1,2}^* = 0$ as an approximation. If $d_{1,2}^* = 0$ on the interval then $N'(d_{1,2}^*) = \frac{1}{\sqrt{2\pi}}$ and $N(d_{1,2}^*) = \frac{1}{2}$. This together with the fact that $\int \frac{e^{-\beta(\tau-u)}}{\sqrt{\tau-u}} \mathrm{d}u = -\sqrt{\frac{\pi}{\beta}} \mathrm{erf}(\sqrt{\beta(\tau-u)})$, where erf is the error function, we can approximate the second part of the integral, and the mixed derivative quite accurately.

Step 3: Evolve $c(\tau)$. After the mixed derivative is calculated we hold it constant for one step and evolve the boundary one step using equation (2.6). Since we hold the mixed derivative constant for one step we see that equation (2.6) becomes an ODE and we can evolve it using any ODE solver. In numerical examples we use the explicit RK2 method. Once we evolve the boundary one step we repeat Steps 2 and 3 until we know the boundary at the desired time to expiration.

Step 4: Calculate the price of the put. Once we know the boundary for all values between 0 and τ we can use equation (2.14) to find the value of an American put at any value of x. Again we need to use numerical integration but this time the integral is very simple because the integrand does not blow up, so we can use any standard numerical integration technique. Numerical results are presented in the next section.

2.5. Numerical Results

To compare the speed and accuracy we compute the (long-dated) option values over the set of parameters presented on Table 3a in AitSahlia and Carr (1997). Here we assume that the underlying asset is a constant dividend paying stock and thus $b = r - \delta$, where δ is the dividend yield and we use $\hat{x} = 6.5q$. The value of the put is calculated when $x = 80$, 85, 90, 95, 100, 105, 110, 115, 120 for the parameter values $q = 100$, $\tau = 3$, $\sigma = 0.4$, $r = 0.06$, and $\delta = 0.02$. Then holding all other parameters fixed at this level we evaluate the at-the-money put with the parameters $r = 0.02, 0.04, 0.08, 0.1$, $\delta = 0, 0.04$, $\sigma = 0.3, 0.35, 0.45, 0.5$, and $\tau = 0.5, 1, 1.5, 2, 2.5, 3.5, 4, 4.5, 5, 5.5$. This leads to 21 sets of parameters where we evaluate the American put.

We compare these values to the values calculated using the very accurate, yet very slow, binomial tree method. We then compare this accuracy measure to the accuracy of four other computational methods: the finite difference moving boundary method in Muthuraman (2008), the Brennan–Schwartz method, the front fixing method in Nielsen et al. (2002) and the standard integral method in Carr et al. (1992). A more comprehensive comparison of other numerical methods can be found in Muthuraman (2008). The Brennan–Schwartz and the moving boundary method have some similarity to the static grid and dynamic grid methods since they find the boundary and evolve the price by time stepping. However in these methods the boundary is always considered to be at a grid point and the way it is found, by evolving equation (2.1) over a large domain several times, is much slower than our method, evolving an ODE. We compare to the standard integral method because we have created a modified integral method that uses the boundary evolution equation and we would like to see if this is advantageous. The front fixing method is considered too, because it also removes the moving boundary by a change of variables similar to the one considered here, however this method is very slow and inaccurate because it must solve a large system of nonlinear equations at each time step.

The measure of accuracy here is the same as the one used by Broadie and Detemple (1997), root mean squared relative error, RMSE, and we consider the "exact" price to be the average of a 10,000 and a 10,001 step binomial tree approximation, as in AitSahlia and Carr (1997). RMSE is defined as RMSE $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{\text{approx}_i - \text{exact}_i}{\text{exact}_i})^2}$ where the sum

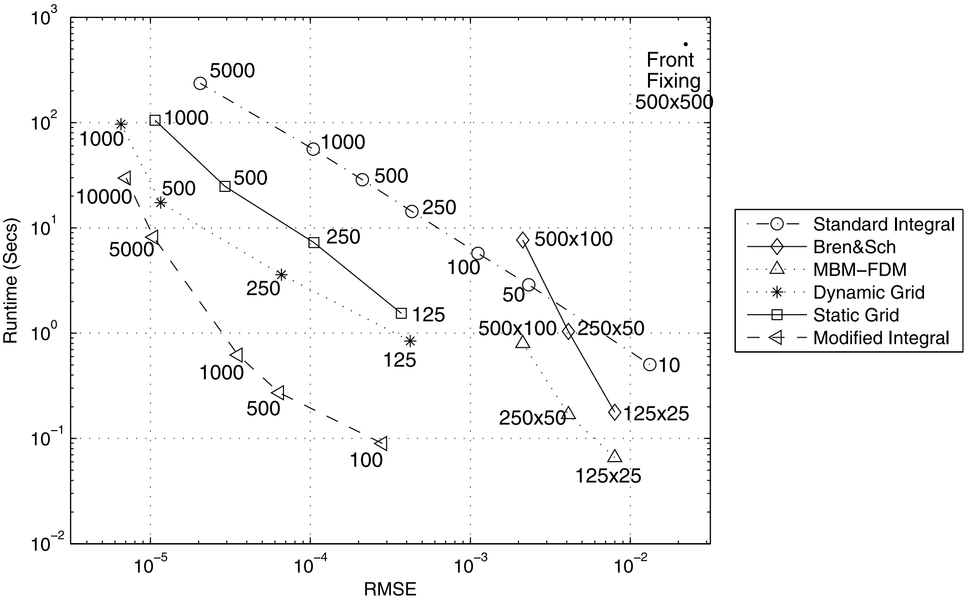


FIGURE 2.6. RMSE vs. runtime for constant volatility.

is taken over all numerical experiments, approx_i is the value of the ith put found by the approximate numerical method, and exact_i is the "exact" value of the put.

The measure of speed is simply average total computational time. We calculate the speed and error of these methods over several grid sizes and show the results in Figure 2.6. For the dynamic and static grid methods the labels refer to the number of spacial grid points; the number of grid points in time to expiry is determined by the CFL condition: the step size in τ is proportional to the square of the step size in x, which guarantees that the matrices used for evolution are positive definite. It is important that the evolution matrices be positive definite because if values on the main diagonal are negative then roundoff error can accumulate quickly, see Courant, Friedrichs, and Lewy (1967). For the modified and standard integral methods the labels refer to the number of time grid points, and for the Brennan–Schwartz, moving boundary and front fixing methods the labels refer to the number of space and time grid points. All computations were performed in Matlab on a PC with a 3.06 GHz processor and 4GB of RAM running Ubuntu Linux 10.10.

All analysis here was performed with $0 \leq \delta \leq r$. Unfortunately, when $r < \delta$ it can happen that our initial approximation of $c(\tau_o)$ is greater than $\frac{r}{\delta}q$. This means that the denominator of equation (2.6) is negative and the whole equation is positive, indicating that the boundary is increasing in time-to-expiry, which is clearly incorrect. To overcome this problem we can use other methods to initialize the American put, such as a few steps in the integral method or any variety of short-time asymptotic approximations. It seems that using a few steps in the integral method is favorable to short-time asymptotic approximations because the time required to initialize with the integral method does not increase total computational time by much and it typically results in less error than short-time asymptotic approximations. Alternatively if $\delta < 0$, (i.e., $b > r$), there is no change to the method and speed and accuracy are comparable to the existing results.

We can see that the static and dynamic grid methods perform better than the standard integral method in both computational time and accuracy. They also provide better accuracy than the Brennan and Schwartz method and the moving boundary method. We also see that the front fixing method is the worst method considered, as is also seen in Muthuraman (2008).

The dynamic grid method is faster than the static grid method despite requiring more time steps because one time step of an explicit method, used in the dynamic grid method, can be faster than one time step of an implicit method, used in a static grid method. The dynamic grid method forces us to use an explicit method because of the nonlinearities. Each step in this explicit method results in a few matrix multiplications (depending on the order of the Runge–Kutta method) whereas the Crank–Nicolson method requires matrix multiplication and factorization, to solve a system of equations, at each time step. It is not possible to prestore the matrix factorization before evolving the system because at each step the matrix changes and thus the factorization changes as well. Even though the dynamic grid method takes more time steps than the static grid method, each step in the dynamic grid method is faster than a step in the static grid method and this trade off comes out in favor of the dynamic grid method for most mesh sizes.

Even more impressive than the static and dynamic grid methods is the modified integral method. The modified integral method offers a huge improvement over the standard integral method in both computational time and accuracy. It also out performs the static and dynamic grid methods, especially on a coarse grid. We do not directly extend the modified integral method to stochastic volatility but this would be an interesting direction for future research following Detemple and Tian (2002).

We would also like to know how error in these methods depends on grid size. To do this we will perform two convergence studies where we systematically decrease the step size in the x and τ variables. The first study will be performed on the Fixed Grid and Dynamic Grid methods. In this study when we reduce the step size in x (or ω) linearly, we reduce the step size in τ quadratically, to maintain the CFL condition. We then calculate the L^2 error and find the slope of the log step size versus the log error, this gives us the order of accuracy of the method.

To approximate the L^2 error we compute the price of the put for $x \in [80, 120]$ at $\tau = 3$ for the first parameter set described above. When we perform this regression we find the slope is 1.985 for the dynamic grid method and 1.419 for the fixed grid method, suggesting that the dynamic grid method is second-order accurate. The most likely reason that the fixed grid method loses some accuracy is the nonuniform grid spacing at the boundary; the small distance between the boundary and the first grid point can dominate finite difference calculations.

In the second study we examine the effect of $\Delta\tau$ on error. We perform this test only on the modified integral method. Here we systematically decrease the step size in τ, there is no step size in x, and again approximate the L^2 error over the same domain as in the previous example. When we perform this regression we find the slope is 0.949. This method is only first-order accurate, despite using what seemed to be a second-order finite difference method for $c(\tau)$, because the calculation of the mixed derivative at the boundary, which is only first-order accurate in time, dominates the error.

3. STOCHASTIC VOLATILITY

In this section, we seek the boundary evolution equation that characterizes the early exercise boundary when the dynamics of the underlying asset are modeled by a stochastic volatility process. We will also leverage on the derived equation to create a fast and accurate numerical method to approximate the price of an American option. This time however we will only be able to implement the numerical method on a dynamic grid due to grid effects that will be explained later.

Working with stochastic volatility makes pricing options challenging since there are two space dimensions and one time dimension. The space dimensions are x, which represents the price of the underlying asset, and y, which represents the volatility, or some function of the volatility, of the underlying asset.

Unlike the constant volatility case, when we consider stochastic volatility there are several models in literature for the underlying dynamics of the asset's volatility. The popular models are the Heston (1993) model, the Hull and White (1987) model, the Scott (1987) model, and the Stein and Stein (1991) model. As each model uses a different stochastic process for volatility the PDE describing the risk neutral expectation is different for each model. In each of these the authors have worked on pricing European-style derivatives. Of the above models, the Heston model is the most popular and in the next sections we focus mostly on this model but also present some results for the other models above.

3.1. The Boundary Equation

As in the constant volatility case we must partition the computational domain into two distinct regions separated by the early exercise boundary. Now, however, the early exercise boundary is not just a function of time, but also the volatility level because different levels of volatility will lead to different optimal exercise policies. Before we can derive a PDE for the early exercise boundary we must first understand stochastic volatility models. We will begin working with a set of stochastic differential equations that are sufficiently general to accommodate the popular stochastic volatility models. The SDEs are

$$\begin{aligned} dX_t &= \mu X_t dt + f(Y_t) X_t dW_1, \\ dY_t &= \eta(Y_t) dt + \lambda(Y_t) dW_2, \\ \langle dW_1, dW_2 \rangle &= \rho dt. \end{aligned}$$

Here X_t is the stochastic process representing the price of the underlying asset, Y_t represents the volatility of the underlying asset, f, η, and λ are model specific functions, and ρ is the correlation between the two Brownian motions, W_1 and W_2. With these SDE's we can use a dynamic programming argument with Itô calculus and the no-arbitrage argument to write a PDE and boundary conditions that the value of the American put must satisfy in the nonexercise region of the domain, $\{(x, y, \tau)\colon \forall y, \tau \geq 0, x > c(y, \tau)\}$. Here we do not consider dividends for simplicity. The differential equation is

$$\frac{\partial p}{\partial \tau} = \frac{1}{2} x^2 f(y)^2 \frac{\partial^2 p}{\partial x^2} + \frac{1}{2} \lambda(y)^2 \frac{\partial^2 p}{\partial y^2} + \rho \lambda(y) f(y) x \frac{\partial^2 p}{\partial x \partial y} + r x \frac{\partial p}{\partial x} + \eta(y) \frac{\partial p}{\partial y} - rp, \tag{3.1}$$

with boundary conditions

$$p(c(y, \tau), y, \tau) = q - c(y, \tau), \tag{3.2}$$

$$\frac{\partial}{\partial x} p(c(y, \tau), y, \tau) = -1, \tag{3.3}$$

$$\frac{\partial}{\partial y} p(c(y,\tau), y, \tau) = 0, \quad \text{and} \tag{3.4}$$
$$p(x, y, 0) = \max(q - x, 0).$$

We also assume

$$\lim_{y\to\infty} \frac{\partial p}{\partial y} = 0, \tag{3.5}$$

which implies that $\lim_{y\to\infty} \frac{\partial c}{\partial y} = 0$. And for large x we use the same boundary condition as equation (2.5). Equations (3.3) and (3.4) are the smooth pasting conditions for stochastic volatility, as found in Fouque, Papanicolaou, and Sircar (2000). Now that we have the boundary conditions we next seek a differential equation that governs $c(y, \tau)$. We give the proof for the general formulation and later present the results for several specific models.

THEOREM 3.1. *If $\frac{\partial c}{\partial \tau} \neq 0$ and $c(y, \tau)$ is sufficiently smooth, the differential equation that governs $c(y, \tau)$ is*

$$\frac{\partial c}{\partial \tau} = -\frac{\partial^2 p(c, y, \tau)}{\partial x \partial \tau} \frac{1}{2rq} \left(f(y)^2 c^2 - 2\rho\lambda(y) f(y) c \frac{\partial c}{\partial y} + \lambda(y)^2 \left(\frac{\partial c}{\partial y}\right)^2 \right). \tag{3.6}$$

Proof. Similar to the derivation of equation (2.7), in stochastic volatility we also have

$$\frac{\partial}{\partial \tau} p(c(y,\tau), y, \tau) = 0. \tag{3.7}$$

Next we differentiate (3.3) with respect to y and τ, and (3.4) with respect to y giving us

$$\frac{\partial^2 p}{\partial x^2} \frac{\partial c}{\partial y} + \frac{\partial^2 p}{\partial x \partial y} = 0, \tag{3.8}$$

$$\frac{\partial^2 p}{\partial x^2} \frac{\partial c}{\partial \tau} + \frac{\partial^2 p}{\partial x \partial \tau} = 0, \tag{3.9}$$

and

$$\frac{\partial^2 p}{\partial x \partial y} \frac{\partial c}{\partial y} + \frac{\partial^2 p}{\partial y^2} = 0. \tag{3.10}$$

Now combine (3.8) and (3.10) to see that

$$\frac{\partial^2 p}{\partial y^2} = \frac{\partial^2 p}{\partial x^2} \left(\frac{\partial c}{\partial y}\right)^2. \tag{3.11}$$

Next evaluate (3.1) at the boundary and substitute in (3.2), (3.3), (3.4), (3.8), and (3.11), which gives us

$$0 = \frac{1}{2} f(y)^2 c^2 \frac{\partial^2 p}{\partial x^2} - \rho\lambda(y) f(y) c \frac{\partial^2 p}{\partial x^2} \frac{\partial c}{\partial y} + \frac{1}{2}\lambda(y)^2 \left(\frac{\partial c}{\partial y}\right)^2 \frac{\partial^2 p}{\partial x^2} - rq. \tag{3.12}$$

Finally plug in (3.9) to equation (3.12) and rearrange terms to obtain the desired result, (3.6). $\square$

Now that we have the formula for the general stochastic volatility formulation we plug in the model specific functions, f, η, and λ, and describe the boundary equations for the four models above in Table 3.1. In all of these models the market price of risk is assumed to be zero but it could be inserted into the differential equations without much effort because the coefficient of the first derivative in y, which is where the market price of risk enters the system, is not present in the boundary evolution equation.

In the statement of Theorem 3.1 we only derive the boundary evolution equation when $\frac{\partial c}{\partial \tau} \neq 0$ for all values of y. This guarantees that we do not divide by zero when plugging equation (3.9) into (3.12). If this is not true, then the boundary just does not move at that point. It seems however for the Heston model and the Hull and White model that as $y \to 0$ we also have $\frac{\partial c}{\partial \tau} \to 0$ for all values of τ. This would mean that $c(0, \tau) = q$ and $p(x, 0, \tau) = 0$ for all τ and $x \geq q$. For the Hull and White model this is not surprising because the variance in this model follows a Geometric Brownian Motion, which stays at zero forever if the process is ever zero, almost surely. This means that the value of the underlying becomes deterministic and thus an out-of-the-money put can have no value when $y = 0$, which can be used as a boundary condition for the Hull and White model.

This point is subtle because even though a Geometric Brownian Motion can never reach zero, if it starts at a positive value, the PDE for the value function needs a boundary condition. The boundary condition chosen here needs to agree with the dynamics of the stochastic process, and since a Geometric Brownian Motion that starts at zero must stay at zero, this is the boundary condition that we must use.

The above economic reasoning, however, does not make sense for the Heston model because the variance follows a square root process which becomes positive immediately after hitting zero, almost surely (for certain parameter values satisfying the Feller Condition, zero is inaccessible to the variance process, like Geometric Brownian Motion, but we still need a boundary condition.) This means that the value of the underlying cannot be deterministic and thus an out-of-the-money put must have positive value when $y = 0$, implying that $\lim_{y\to 0^+} \frac{\partial c}{\partial y} = -\infty$.

In this case the rate that the derivative explodes must be very specific. It must go to infinity like $\frac{-1}{\sqrt{y}}$. If it goes to infinity any faster then the last term in the Heston boundary equation will go to infinity and so will the whole boundary equation. If it goes to infinity any slower then the last term will go to zero and so will the whole boundary equation. If the derivative does go to infinity at the right speed then the last term becomes indeterminate, which makes

$$\frac{\partial c(0, \tau)}{\partial \tau} = -\frac{\partial^2 p(c, y, \tau)}{\partial x \partial \tau} \frac{v^2}{2rq} \gamma. \tag{3.13}$$

Although the above constant, γ, is unknown we can interpret this as a boundary condition for the Heston model, which will be explained in further detail in the next section.

Figure 3.1 shows the state space partitioned to the exercise region and the continuation region for the Heston model. The two regions are separated by the early exercise surface. The exercise region is below the surface and the price of the put is equal to its intrinsic value there. The continuation region is above the surface and there the price of the put is governed by equation (3.1).

TABLE 3.1.
Boundary Evolution Equations for Several Popular Stochastic Volatility Models

Model	Stochastic Process	Boundary Equation
Heston	$dX_t = \mu X_t dt + \sqrt{Y_t} X_t dW_1$	$\frac{\partial c}{\partial \tau} = -\frac{\partial^2 p(c,y,\tau)}{\partial x \partial \tau} \frac{1}{2rq}(yc^2 - 2\rho v y c \frac{\partial c}{\partial y} + v^2 y (\frac{\partial c}{\partial y})^2)$
$\rho \neq 0$	$dY_t = \kappa(m' - Y_t)dt + v\sqrt{Y_t} dW_2$	
Hull & White	$dX_t = \mu X_t dt + \sqrt{Y_t} X_t dW_1$	$\frac{\partial c}{\partial \tau} = -\frac{\partial^2 p(c,y,\tau)}{\partial x \partial \tau} \frac{1}{2rq}(yc^2 + \alpha_2^2 y^2 (\frac{\partial c}{\partial y})^2)$
$\rho = 0$	$dY_t = \alpha_1 Y_t dt + \alpha_2 Y_t dW_2$	
Scott	$dX_t = \mu X_t dt + e^{Y_t} X_t dW_1$	$\frac{\partial c}{\partial \tau} = -\frac{\partial^2 p(c,y,\tau)}{\partial x \partial \tau} \frac{1}{2rq}(e^{2y} c^2 + \beta^2 (\frac{\partial c}{\partial y})^2)$
$\rho = 0$	$dY_t = \alpha(m - Y_t)dt + \beta dW_2$	
Stein & Stein	$dX_t = \mu X_t dt + \vert Y_t \vert X_t dW_1$	$\frac{\partial c}{\partial \tau} = -\frac{\partial^2 p(c,y,\tau)}{\partial x \partial \tau} \frac{1}{2rq}(y^2 c^2 + \beta^2 (\frac{\partial c}{\partial y})^2)$
$\rho = 0$	$dY_t = \alpha(m - Y_t)dt + \beta dW_2$	

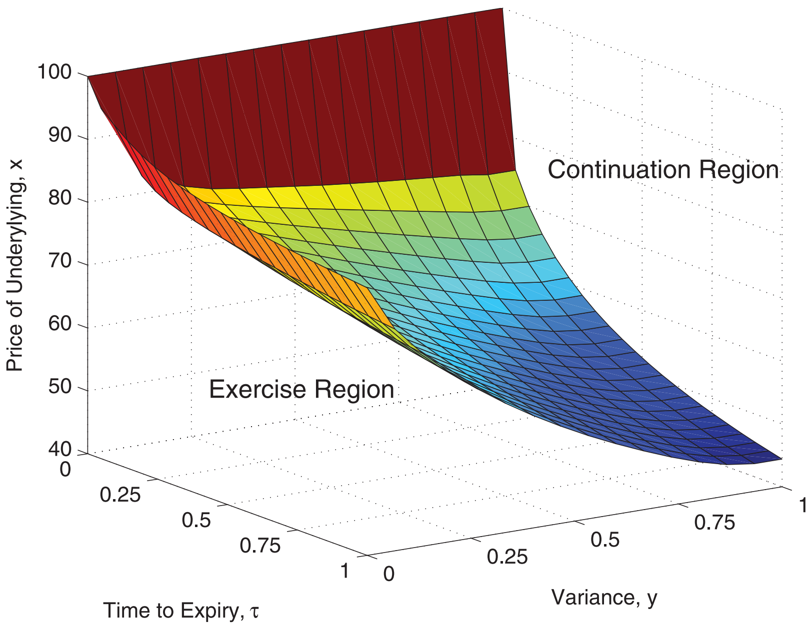


FIGURE 3.1. Partitioned state space for Heston model.

3.2. Numerical Method on a Dynamic Grid

In this section we will focus only on the Heston model of stochastic volatility. We want to transform the no exercise region to a simpler domain that allows for standard finite difference methods. We transform the domain $\{(x, y, \tau)\colon \forall y, \tau \geq 0, x > c(y, \tau)\}$ to $\mathbb{R}^3_+$. The change of variable we use to transform our domain is

$$\begin{aligned} \omega &= x - c(y, \tau) \\ g(\omega, y, \tau) &= p(x, y, \tau). \end{aligned} \tag{3.14}$$

This change of variable involves derivatives in the y variable. The second derivative and the mixed derivative lead to several nonlinear terms in the resulting PDE. Equation (3.1) for the Heston model is

$$\frac{\partial p}{\partial \tau} = \frac{1}{2} y x^2 \frac{\partial^2 p}{\partial x^2} + \frac{1}{2} v^2 y \frac{\partial^2 p}{\partial y^2} + \rho v y x \frac{\partial^2 p}{\partial x \partial y} + r x \frac{\partial p}{\partial x} + \kappa (m' - y) \frac{\partial p}{\partial y} - r p. \tag{3.15}$$

We use the chain rule to find the pricing equation for the g function and the corresponding boundary equation, which are

(3.16)
$$\begin{aligned} \frac{\partial g}{\partial \tau} &= \frac{1}{2} y(\omega + c(y,\tau))^2 \frac{\partial^2 g}{\partial \omega^2} + \frac{1}{2} v^2 y \left(\frac{\partial^2 g}{\partial \omega^2} \left(\frac{\partial c}{\partial y} \right)^2 - \frac{\partial g}{\partial \omega} \frac{\partial^2 c}{\partial y^2} - 2 \frac{\partial c}{\partial y} \frac{\partial^2 g}{\partial \omega \partial y} + \frac{\partial^2 g}{\partial y^2} \right) \\ &\quad + \rho v y (\omega + c(y,\tau)) \left(\frac{\partial^2 g}{\partial \omega \partial y} - \frac{\partial c}{\partial y} \frac{\partial^2 g}{\partial \omega^2} \right) + r(\omega + c(y,\tau)) \frac{\partial g}{\partial \omega} + \kappa (m' - y) \\ &\quad \times \left(\frac{\partial g}{\partial y} - \frac{\partial g}{\partial \omega} \frac{\partial c}{\partial y} \right) - r g + \frac{\partial g}{\partial \omega} \frac{\partial c}{\partial \tau} \end{aligned}$$

and

$$\frac{\partial c}{\partial \tau} = -\frac{\partial^2 g(0, y, \tau)}{\partial \omega \partial \tau} \frac{1}{2rq} \left(yc^2 - 2\rho v y c \frac{\partial c}{\partial y} + v^2 y \left(\frac{\partial c}{\partial y} \right)^2 \right). \tag{3.17}$$

Given this change of variables we seek a numerical method that exploits the equations for boundary and price evolution. The method presented here can be summarized in a three-step process.

Step 1: Initialization. Similar to the constant volatility case, we cannot start the numerical method with the initial conditions when $\tau = 0$, and as such we need to approximate the value of the American option a short time before expiry as a European option. There are two ways to approximate the value of the European option. First we could use the semi closed form solution to European options under the Heston model to find the price of the European a short time before expiry, the details of which can be found in Heston (1993) or Gatheral (2006). Alternatively we could use the constant volatility Black–Scholes equation to find the value of the put a short time before expiration. It might seem that this simplistic method would lead to large error, but it turns out that the two methods have comparable accuracy and the second is significantly faster than the first. The reason is that the functions being integrated in the solution to the European put under the Heston model are highly oscillatory and are dampened very slowly for small values of τ. This makes approximating this integral a very slow process because a large integration domain is required with a fine integration mesh, and so for numerical tests we simply use the Black–Scholes equation to initialize p.

To initialize we need to divide the y domain uniformly between 0 and $\hat{y}$, where $\hat{y}$ is the maximal value of the computational domain. Here again the value of $\hat{y}$ needs to be large enough so that the boundary condition in Equation (3.5) is approximately true for all values of x. At each grid point in y we perform a binary search to find the intersection of the value of the European option and the intrinsic value of the option as in Figure 2.2. If we use the Black–Scholes formula to get the value of the European then we need to set the variance equal to the y grid value. We initialize the boundary at each y grid point as the location of the intersection. Then for each value of y we find the value of the European at n equally spaced grid points, in x, larger than the boundary, where n is chosen large enough so that the boundary condition (2.5) is satisfied. We find that the initial boundary is deceasing in y and as such the maximal value of x for each value of y is also decreasing. After we find the price of the European at all of these grid points we transform the domain using equation (3.14). Figure 3.2 shows how the computational domain looks before the transformation.

Step 2: Calculate $\frac{\partial^2 g(0,y,\tau)}{\partial \omega \partial \tau}$. As in the constant volatility case the hardest part of the algorithm is finding the mixed derivative at the boundary. In the constant volatility case we discussed two ways to calculate this derivative, one in the fixed grid method and one in the dynamic grid method. Here since we only work on a dynamic grid we simply evolve a few grid points greater than the boundary for every value of y according to equation (3.16) without the last term, the grid speed term, using an explicit Runge–Kutta method. We use a standard one-sided finite difference method to calculate the value of the x derivative at the boundary for every value of y after this partial evolution. Then using this value with equation (3.3) we can approximate the value of the mixed derivative by using a first-order finite difference method.

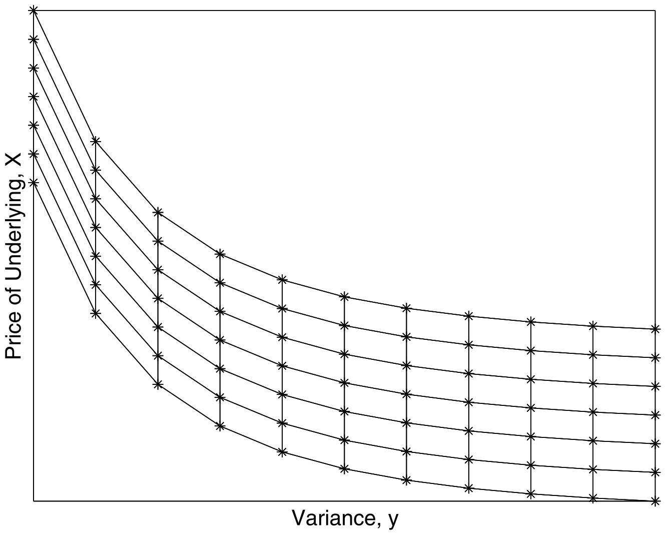


FIGURE 3.2. Computational grid before transformation.

Step 3: Evolve *c* and *g* in time to expiry. As opposed to the method used for constant volatility, we linearize equation (3.16) so that we can use an implicit method to step backwards in time, which dramatically reduces the number of steps required in time to expiry when compared to an explicit method. In constant volatility we could have also linearized the price evolution equation in the dynamic grid section to use an implicit method. However on a fine grid linearization accounts for a large portion of the numerical error and so we only use an explicit method. In stochastic volatility it is not practical to use a fine grid because there are two space dimensions which greatly increases the total number of grid points and therefore we linearize equation (3.16).

We see that in equation (3.16) all the nonlinearities come from multiplying derivatives of g with derivatives of c. This means that if we can approximate the derivatives of c then we can use them to linearize the evolution equation for g. To linearize this equation we must get the first- and second-order derivatives of the boundary with respect to y. To do this we simply evolve the boundary one step using a Runge–Kutta method and compute the derivatives for the boundary at the τ and $(\tau + 1)^{\text{st}}$ steps using standard finite difference methods. Then using the values computed here we plug them into Crank–Nicolson matrices A and B, where A and B are block tridiagonal matrices satisfying the equation $A \cdot g^{\tau} = B \cdot g^{\tau+1}$. Here we plug the values of the derivatives before the evolution into the A matrix and the values of the derivatives after the evolution into the B matrix. One important fact to remember is that at each step the matrices A and B must be recalculated because the boundary and the derivatives of the boundary have changed. After we evolve the price function we let the boundary be equal to the value at the $(\tau + 1)^{\text{st}}$ step.

There is still a boundary condition we need to address: the boundary when y goes to 0. For this we simply assume that the constant in equation (3.13) is attained at the second y grid point and that the value of p evolves with the standard PDE when 0 is inserted for y, which eliminates several terms. After we evolve g one step we repeat Steps 2 and 3 until we reach the desired time to expiration and change the variables back to x and p.

We cannot adopt this method onto a fixed grid, as we did with constant volatility, because the boundary is decreasing in y. Say for a specific y value the boundary is between the 99th and 100th x grid values and for the next y value the boundary is between the 98th and 99th x grid values. We will then get a discontinuity in the calculation of the mixed derivative when this happens. This discontinuity in the mixed derivative leads to a discontinuous boundary which in turn leads to large error in the price of the put and as such we need a dynamical grid method. The boundary is also decreasing in time to expiry and so this phenomenon could also occur in the τ variable. The effect, however, is less drastic in τ than in y because at each discrete step in τ we numerically approximate derivatives in y, whereas we relate derivatives in τ to derivatives in y using equation (3.16) removing the need for continuous derivatives in τ. This relationship is why we were able to use a static grid for constant volatility.

Also, as in the constant volatility case we add extra grid points to each value of y every time that the boundary decreases below a certain value. This again has the benefit of maintaining accuracy for options that are out of the money.

3.3. Numerical Results

Numerical comparison of speed and accuracy is more challenging for stochastic volatility than for constant volatility because finding a "true" price for the option is not clear. In this section we only compute the price of the put for eight set of parameters, the "true" values were calculated by Jari Toivanen using his component-wise splitting method on a very fine mesh. The value of the put is calculated when $x = 8, 9, 10, 11, 12$, $y = 0.0625, 0.25$, and $\tau = 0.25$ for the parameter values $r = 0.1$, $v = 0.9$, $\kappa = 5$, $m' = 0.16$, and $\rho = 0.1$. Then holding all other parameters fixed we evaluate the puts with the parameters $r = 0.08, 0.12$, $v = 0.7, 1.1$, $\kappa = 2.5$, and $\rho = 0.05, 0.15$.

We compare our method to two existing methods, the PSOR and the moving boundary method, MBM, presented in Choklingham and Muthuraman (2011), in Figure 3.3. We only compare our method against these methods because although the PSOR method is quite slow, Ikonen and Toivanen (2008) find that it is the simplest to implement, and in Chockalingham and Muthuraman (2011) the authors find that the moving boundary method was the fastest method tested. As in the constant volatility case we plot the root mean squared relative error of the methods versus the total computational time. For the moving boundary method and the PSOR the labels refer to the number of x, y, and τ grid points. For the dynamic grid method the labels refer to the number of x and y grid points, and the number of time grid points is determined by the CFL condition.

The nonlinearities of the dynamic grid method unfortunately cause the necessary number of steps in time to expiry to be quite large, despite the linearization of equation (3.16), this however is offset by the speed with which each time step is executed versus the moving boundary method and the PSOR. Both of these methods must search for the early exercise boundary while our method knows exactly where it is. We see that for the coarsest grid the moving boundary method is slightly better than our method, however on finer grids our method performs significantly better. For the finest grid our method is almost three times faster than the moving boundary method.

4. CONCLUDING REMARKS

Boundary evolution equations have significant computational benefit when one relies on dynamic grids that are evolved with the boundary during the solution process. The key insight into the construction of efficient numerical methods is that we do not have to iteratively guess the location of the boundary at each step, rather the boundary evolution

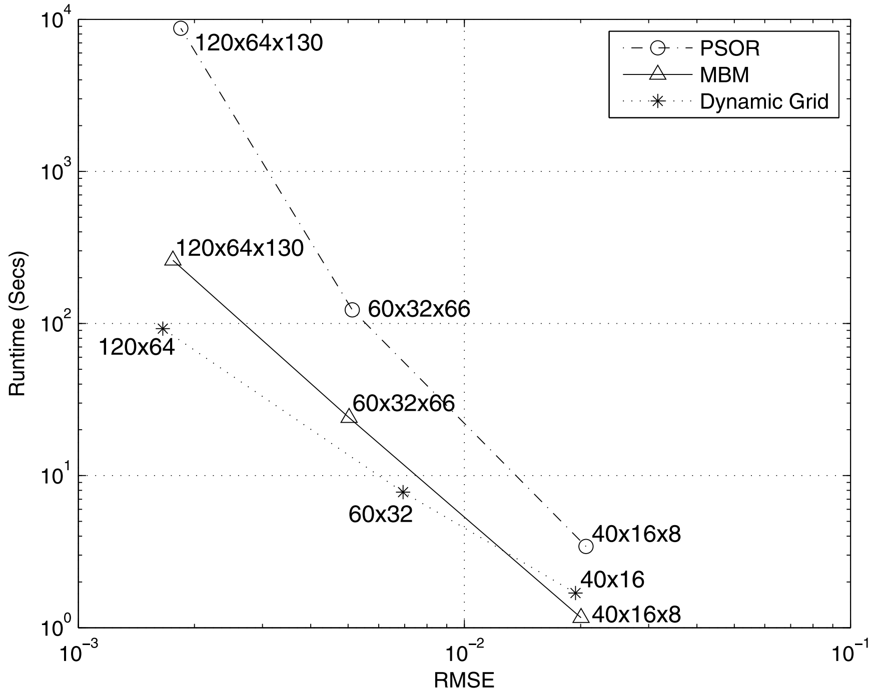


FIGURE 3.3. RMSE vs. runtime for stochastic volatility.

equation tells us its location. Moreover, by evolving the grid along with the boundary one gets the added benefit of minimizing the error in approximating the boundaries with a predefined grid structure.

The American option pricing problems studied here belong to the much larger class of optimal stopping problems in stochastic control. Most optimal stopping problems do not have analytical solutions and are difficult to solve, especially when the complexity of the state evolution equation increases. In many cases the location of the boundary that separates the stopping and continuation regions is of primary interest. As such boundary evolution equations can provide insight into the structure and nature of these boundaries. The derivation of the boundary evolution equations rely on the smooth pasting condition at the interface between the stopping and continuation regions. Similar smooth pasting conditions are also common in several derivative securities and other optimal stopping problems, such as simultaneous hypothesis testing and earliest detection problems. See Peskir and Shiryaev (2006) for examples of other optimal stopping problems.

In the Black–Scholes setting we presented a modified integral method for pricing American options that relied on an integral representation of the price of the American option. This method proved to be extremely fast and accurate in the simple case of Black–Scholes. An extension to multifactor models of the integral representation has been presented in Detemple and Tian (2002) and an interesting direction of future research would be to apply the boundary evolution equations for stochastic volatility found in this paper to a modified integral method for multifactor models.

Two other classes of stochastic control problems whose solutions are characterized by free-boundary problems are singular and impulse control. In these problems the state process is not terminated at the boundary, but a control is applied to it. Both deriving boundary evolution equations and constructing computational methods for these would be interesting future work. The ideas in this paper cannot be immediately extended to optimal stopping problems with multiple boundaries and this would be interesting future work as well.

REFERENCES

AITSAHLIA, F., and P. CARR (1997): American Options: A Comparison of Numerical Methods, in *Numerical Methods in Finance*, L. Rogers and D. Talay, eds. Cambridge, UK: Cambridge University Press, pp. 67–87.

BLACK, F., and M. SCHOLES (1973): The Pricing of Options and Corporate Liabilities, *J. Polit. Econ.* 81, 637–654.

BRENNAN, M., and E. SCHWARTZ (1977): The Valuation of American Put Options, *J. Finance* 32, 449–642.

BROADIE, M., and J. DETEMPLE (1996): American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods, *Rev. Finan. Stud.* 9, 1211–1250.

BROADIE, M., and J. DETEMPLE (1997): Recent Advances in Numerical Methods for Pricing Derivative Securities, in *Numerical Methods in Finance*, L. Rogers and D. Talay, eds. Cambridge, UK: Cambridge University Press, pp. 43–66.

BROADIE, M., J. DETEMPLE, E. GHYSELS, and O. TORRES (2000): American Options with Stochastic Dividends and Volatility: A Nonparametric Investigation, *J. Econometr.* 94, 53–92.

BROADIE, M., and P. GLASSERMAN (1997): Pricing American-Style Securities by Simulation, *J. Econ. Dyn. Contl.* 21, 1323–1352.

BUNCH, D., and H. JOHNSON (2002): The American Put Option and Its Critical Stock Price, *J. Finance* 55, 219–237.

CARR, P., R. JARROW, and R. MYNENI (1992): Alternative Characterizations of American Put Options, *Math. Finance* 2, 87–106.

CHEN, S., B. MERRIMAN, S. OSHER, and P. SMEREKA (1997): A Simple Level Set Method for Solving Stefan Problems, *J. Comput. Phys.* 135, 8–29.

CHEN, X., and J. CHADAM (2007): A Mathematical Analysis of the Optimal Exercise Boundary for American Put Options, *SIAM J. Math. Anal.* 38, 1613–1641.

CHOCKALINGAM, A., and K. MUTHURAMAN (2011): American Options under Stochastic Volatility, *Operat. Res.* 59, 793–804.

CLARKE, N., and K. PARROTT (1999): Multigrid for American Option Pricing with Stochastic Volatility, *Appl. Math. Finance* 6, 177–195.

COURANT, R., K. FRIEDRICHS, and H. LEWY (1967): On the Partial Difference Equations of Mathematical Physics, *IBM J. Res. Develop.* 11, 215–234.

COX, J., S. ROSS, and M. RUBINSTEIN (1979): Option Pricing: A Simplified Approach, *J. Finan. Econ.* 7, 229–263.

CRANK, J., and P. NICOLSON (1947): A Practical Method for Numerical Evaluation of Solutions of Partial Differential Equations of the Heat-Conduction Type, *Proc. Camp. Phil. Soc.* 43, 50–67.

DETEMPLE, J., and W. TIAN (2002): The Valuation of American Options for a Class of Diffusion Processes, *Manage. Sci.* 48, 917–937.

EVANS, J., R. KUSKE, and J. KELLER (2002): American Options on Assets with Dividends Near Expiry, *Math. Finance* 12, 219–237.

FOUQUE, J., G. PAPANICOLAOU, and K. SIRCAR (2000): *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge: Cambridge University Press.

GATHERAL, J. (2006): *The Volatility Surface: A Practitioner's Guide*, Hoboken, NJ: Wiley.

GOODMAN, J., and D. OSTROV (2002): On the Early Exercise Boundary of the American Put Option, *SIAM J. Appl. Math.* 62, 1823–1835.

HAYES, E. (2006): The Application of a Semi-analytical Method for Computing Asymptotic Approximations to Option Prices, PhD thesis, Courant Institute of Mathematical Sciences, New York University.

HESTON, S. (1993): A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Rev. Finan. Stud.* 6, 327–343.

HUANG, J., M. SUBRAHMANYAM, and G. YU (1996): Pricing and Hedging American Options: A Recursive Integration Method, *Rev. Finan. Stud.* 9, 277–300.

HULL, J., and A. WHITE (1987): The Pricing of Options on Assets with Stochastic Volatilities, *J. Finance* 42, 281–300.

IKONEN, S., and J. TOIVANEN (2007): Componentwise Splitting Methods for Pricing American Options Under Stochastic Volatility, *Int. J. Theor. Appl. Finance* 10, 331–361.

IKONEN, S., and J. TOIVANEN (2008): Efficient Numerical Methods for Pricing American Options under Stochastic Volatility, *Numer. Methods Partial Different. Equat.* 24, 104–126.

ISERLES, A. (2008): *A First Course in the Numerical Analysis of Differential Equations*, Cambridge, UK: Cambridge University Press.

JACKA, S. (1991): Optimal Stopping and the American Put, *Math. Finance* 1, 1–14.

KARATZAS, I., and S. SHREVE (1998): *Methods of Mathematical Finance*, New York: Springer.

KIM, I. (1990): The Analytic Valuation of American Options, *Rev. Finan. Stud.* 3, 547–772.

LAWRENCE, P., and S. SALSA (2009): Regularity of the Free Boundary of an American Option on Several Assets, *Comm. Pure Appl. Math.* 62, 969–994.

LONGSTAFF, F., and E. SCHWARTZ (2001): Valuing American Options by Simulation: Simple Least-Squares Approach, *Rev. Finan.Stud.* 14, 113–147.

MERTON, R. (1992): *Continuous-Time Finance*, Oxford: Blackwell.

MUTHURAMAN, K. (2008): A Moving Boundary Approach to American Option Pricing, *J. Econ. Dyn. Contl.* 32, 3520–3537.

MYNENI, R. (1992): The Pricing of the American Option, *Ann. Appl. Probab.* 2, 1–23.

NIELSEN, B., O. SKAVHAUG, and A. TVEITO (2002): Penalty and Front-Fixing Methods for the Numerical Solution of American Option Problems, *J. Comput. Finance* 5, 69–97.

PESKIR, G., and A. SHIRYAEV (2006): *Optimal Stopping and Free-Boundary Problems*, Basel: Birkhäuser Verlag.

SCOTT, L. (1987): Option Pricing When the Variance Changes Randomly: Theory, Estimation, and an Application, *J. Finan. Quantit. Anal.* 22, 419–438.

STAMICAR, R., D. SEVCOVIC, and J. CHADAM (1999): The Early Exercise Boundary for the American Put Near Expiry: Numerical Approximations, *Can. Appl. Math. Quart.* 7, 427–444.

STEIN, E., and J. STEIN (1991): Stock Price Distributions with Stochastic Volatility: An Analytic Approach, *Rev. Finan. Stud.* 4, 727–752.

THOMPSON, J., Z. WARSI, and C. MASTIN (1985): *Numerical Grid Generation: Foundations and Applications*, New York: Elsevier Science.

TILLEY, J. (1993): Valuing American Options in a Path Simulation Model, *Transactions of the Society of Actuaries* 45, 83–104.

VAN MOERBEKE, P. (1975): On Optimal Stopping and Free Boundary Problems, *Arch. Rat. Mech. Anal.* 60, 101–148.

WILMOTT, P. (1998): *Derivatives*, Chichester: Wiley.

WU, L., and Y. KWOK (1997): A Front-Fixing Finite Difference Method for the Valuation of American Options, *J. Finan. Eng.* 6, 83–97.

MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 533–566

PORTFOLIOS OF AMERICAN OPTIONS UNDER GENERAL PREFERENCES: RESULTS AND COUNTEREXAMPLES

VICKY HENDERSON

Oxford-Man Institute, University of Oxford

JIA SUN

China Credit Rating Co.

A. ELIZABETH WHALLEY

Warwick Business School, University of Warwick

We consider the optimal exercise of a portfolio of American call options in an incomplete market. Options are written on a single underlying asset but may have different characteristics of strikes, maturities, and vesting dates. Our motivation is to model the decision faced by an employee who is granted options periodically on the stock of her company, and who is not permitted to trade this stock. The first part of our study considers the optimal exercise of single options. We prove results under minimal assumptions and give several counterexamples where these assumptions fail—describing the shape and nesting properties of the exercise regions. The second part of the study considers portfolios of options with differing characteristics. The main result is that options with comonotonic strike, maturity, and vesting date should be exercised in order of increasing strike. It is true under weak assumptions on preferences and requires no assumptions on prices. Potentially the exercise ordering result can significantly reduce the complexity of computations in a particular example. This is illustrated by solving the resulting dynamic programming problem in a constant absolute risk aversion utility indifference model.

KEY WORDS: American options, utility indifference pricing, employee stock options.

1. INTRODUCTION

Our aim in this paper is to provide results concerning the optimal exercise of American options under a minimal set of assumptions on prices, valuation methodology, and the agent's preferences. The goal is to characterize optimal behavior in terms of exercise ordering for an agent with a portfolio of American calls of differing strikes, maturities, and vesting dates.

The first part of the paper considers single American options and revisits a classic Cox and Rubinstein (1985) result concerning the dependence of the option's continuation

The authors thank David Hobson and participants at seminars at Imperial College London and the A.Ma.Me.F Workshop (Berlin, September 2010) for comments. We also thank an anonymous referee whose comments assisted in clarifying our presentation.

Manuscript received February 2011; final revision received April 2012.

Address correspondence to Vicky Henderson, Oxford-Man Institute, University of Oxford, Eagle House, Walton Well Road, Oxford OX2 6ED, UK; e-mail: vicky.henderson@oxford-man.ox.ac.uk.

DOI: 10.1111/mafi.12008

region upon the strike and option maturity. In particular, a basic result is that if two identical agents each held an American call, the agent with the option with the lower strike/earlier maturity/shorter vesting period would exercise first. Put another way, the continuation region for the option with lower strike/earlier maturity/shorter vesting period is contained within that of the high strike/later maturity/longer vesting period. We call this the "nesting" property. We show this result only requires that agents prefer more to less and prefer cash sooner, and that the valuation mechanism has a monetary property (which we define later). We give several examples where nesting holds (in particular, in a utility indifference model with constant absolute risk aversion [CARA]) and a counterexample (for which we can perform explicit computations) where nesting fails as the valuation is not monetary. On a similar theme, we give conditions under which a higher dividend rate leads to a smaller continuation region. This generalizes results given for exponential Brownian prices in a utility indifference model by Carpenter, Stanton, and Wallace (2010).

We give proofs using probabilistic techniques for each of our results, and we avoid calculations. In particular we use coupling methods, see Lindvall (1992). Indeed, for each of the results described already, we do not need to know anything about the actual shape of the exercise region(s) or boundary(ies) themselves and, as our examples show, these regions may not be simply described by a single threshold separating the continuation and exercise region.

To illustrate how complex exercise boundaries can be, we also study a far more specialized model with risk neutral pricing and diffusion prices. We include these results because some are new and original in their own right, and also they provide a useful comparison to the utility indifference model. We give conditions on the dividend rate such that the exercise region is characterized by a single threshold. A counterexample with state-dependent dividends results in disconnected exercise regions. To obtain a single threshold which is also monotonic in time requires the additional restriction of a time-homogeneous diffusion. This is well known but not always correctly stated (see Kim 1990). A simple counterexample with time-dependent volatility violates the result.

The second part of the paper considers to what extent the results for single American options can be extended to a portfolio of options with differing characteristics. In a complete market, or under pricing by risk-neutral expectation, the consideration of a portfolio of different options is no more difficult than that of a single option. In such a situation, options can be treated independently as the presence of other options does not alter the agent's strategy. However, our interest lies in incomplete markets for an agent with a nonlinear preference structure. Our canonical example of such a market will be the situation of an employee who is restricted from hedging her portfolio of employee stock options and for whom the exercise strategy for a particular option depends upon the rest of her portfolio.

Our main result is that given a portfolio of American call options with a comonotonic set of strikes, maturities, and vesting dates, it is optimal to exercise the options in order of increasing strike. We only require our minimal assumptions on preferences—our agent must prefer more to less and prefer cash sooner—and we do not make any assumptions on the price or on the valuation methodology. The proof relies on a coupling style argument where we consider two agents and show that regardless of the strategy the "suboptimal" agent follows, the "optimal" agent can do better by exercising in order of increasing strike.

Our primary application of the exercise ordering result is to consider a utility indifference pricing model for a portfolio of American options. An introduction to the large literature on utility indifference pricing can be found in Carmona (2008), in particular the survey of Henderson and Hobson (2008). The portfolio exercise problem results in a mixed optimal control and multiple stopping problem reflecting the choices over investment and exercise of multiple options.[1] In general, if there is no comonotonicity, we need to consider all possible exercise orderings. In contrast, if there is comonotonicity, our results show that the problem is reduced to a single dynamic programming exercise. With n distinct options, this is a saving of a factor of $n!$. As an illustration, we solve the dynamic programming problem for CARA and give examples in both in the comonotonic and noncomonotonic cases. This example extends the model of Leung and Sircar (2009) who treat single (and multiple identical) American options under the CARA utility indifference pricing model. This will be discussed further in the next section.

2. EMPLOYEE STOCK OPTIONS

The aim of this section is to give a description of our canonical example and the related literature and briefly discuss the relevance of our results for employee stock options. We consider an employee who has been granted American call options on the stock of her company on a periodic basis. According to the US National Center for Employee Ownership (2008), there exist about 3,000 broad-based employee stock option plans in the United States with about 9 million participants. Commonly these options are granted at-the-money, with maturities of 10 years, and vesting periods of 3 years, although many variations exist. Before the vest date, the option is not permitted to be exercised by the employee. At the vest date, the option becomes American and the employee can exercise it at any time up to and including the expiry date. After a few years, the employee will have accumulated a portfolio of American calls with various strikes, maturities, and vesting dates. We consider the optimal exercise of such an option portfolio—an important step in understanding the cost of granting options from the company perspective—and it is pertinent to consider exercise behavior since it can be observed in practice (unlike say, the subjective value of options to the employee which is unobservable). Note that since both subjective (here, utility indifference) value and the cost to the company are obtained easily once optimal exercise is known, it is sufficient to concentrate on the exercise decision.

Since employees are not permitted to trade in the company stock, and cannot directly sell or transfer their options, they can only unwind their risk exposure by exercising options or perhaps by trading other assets such as a market index. As such, the employee faces an incomplete market. Models based on utility-indifference pricing of American options capture many of the important aspects of the employee's situation. Indeed, there are a number of papers in this vein, notably Carpenter et al. (2010), Grasselli and Henderson (2009), Henderson (2007), Leung and Sircar (2009), and Rogers and Scheinkman (2007), all of whom study American option pricing under utility indifference for single or multiple identical options. (There is also a long literature in finance focussing on one-period or binomial models typically without partial hedging; for a discussion and references, see the survey of Henderson and Sun 2010. The first treatment of employee stock options in a utility indifference framework with partial hedging was Henderson

[1] Several papers consider simulation approaches to multiple exercise problems, see Bender (2011), Ibáñez (2004), and Meinshausen and Hambly (2004) and a Malliavin approach of Carmona and Touzi (2008). These are all concerned with pricing under risk neutral expectation, under specific price models and thus there is no mixed control problem. Additionally, since the main application is swing options they are concerned with exercise constraints which take the form of intervals of time.

2005 who considered European style payoffs.) The above papers make particular choices of utility function (CARA or constant relative risk aversion [CRRA]) and all assume exponential Brownian motions for the stock price and the correlated traded asset, and solve the resulting free boundary problem. This is done explicitly in some special cases (Henderson 2007; Grasselli and Henderson 2009) or via numerical solutions.

Several of our results for single American options are relevant to this literature. We highlight two here. First, we give an example (extended from Carpenter et al. 2010) of preferences where there are several disconnected exercise and continuation regions. This is an important observation from the perspective of employee stock option modeling, as there is a strand of literature (see, e.g., Cvitanic, Wiener, and Zapatero 2008) which assumes an exogenous form for a single boundary of threshold form, rather than deriving the threshold(s) endogenously via a utility (or other) model. It is therefore important to recognize the limitations of such an exogenous specification. Second, we give two alternative sets of conditions under which a higher dividend rate leads to a smaller continuation region. This generalizes results given for exponential Brownian prices in a utility indifference model by Carpenter et al. (2010). In fact, our result is not limited to the utility indifference setting but only requires weak assumptions on preferences.

The main interest of this paper for the employee stock option literature is the treatment of portfolios of American options with different strikes, maturities, and vesting dates. Carpenter et al. (2010) only consider a single American call option in their paper, but acknowledge that in a more realistic portfolio setting "It would be useful to understand which options are most attractive to exercise first...." Although some of the aforementioned papers recognize the need to study multiple options, only options with identical characteristics have been studied. In this case, the employee's risk aversion causes her to unwind risk gradually and thus exercise options intertemporally. For instance, the assumption of infinitely divisible claims (and perpetual options) in Henderson and Hobson (2011) leads to singular control where options are exercised when the price reaches a new maximum, to keep the position below a smooth function.

We consider the exercise of portfolios of American options with differing characteristics. Our main exercise ordering result requires the strikes, maturities, and vest dates are comonotonic. Firms often grant employee stock options which are at-the-money and with a fixed vesting date of 3 years. If such options are granted in a bull market, then their characteristics will indeed be comonotonic. Since the theorem also holds for ordered but random maturities, its conclusion is still true if the employee faces employment termination risk or an exogenous income shock.

We solve the dynamic programming problem for CARA utility and illustrate the resulting exercise boundaries for a portfolio of options. Our examples highlight the influence of other options on the exercise of a particular option in the portfolio. For example, the existence of a longer dated, higher strike option will cause the agent to exercise a shorter dated, lower strike option earlier. This is in contrast to a risk neutral pricing model under which options can be treated independently and thus do not interact. This extends the model of Leung and Sircar (2009) who solve numerically the dynamic programming problem for single American options under the CARA utility indifference pricing model.

3. EXERCISING SINGLE AMERICAN CALL OPTIONS

We work on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_s\}_{s\geq t}, \mathbb{P})$ and let t denote current time and $(X_s)_{s\geq t}$ the price process of an asset. The market consists of the risky asset X and riskless bonds. We consider an agent who holds a portfolio of American call options on X. A call option on X has strike K, maturity T, and vesting date $V \leq T$. During the vesting period $t \leq s \leq V$ the agent is not permitted to exercise the option. The agent's portfolio will contain American calls of potentially differing strikes, maturities, and vesting dates. We will consider n options with characteristics $(K^{(i)}, T^{(i)}, V^{(i)})_{1 \leq i \leq n}$ with vest dates such that $t \vee V^{(i)} \leq T^{(i)}$. In this section, we develop notation and assumptions to state results for a single American option. We return to the portfolio problem in Section 6.

Let $(I_s^{(t)})_{s \geq t} = (I_s^{(t)}(\omega))_{s \geq t}$ be an increasing process representing the income or cashflow generated by the exercise of options from time t onwards. For a single American call option exercised at stopping time $\tau = \tau(\omega)$ with $t \vee V \leq \tau \leq T$ we have

$$I_s^{(t)}(\tau) := I_s^{(t)}(\tau(\omega)) = (X_{\tau(\omega)} - K)^+ 1_{\{s \geq \tau(\omega)\}}. \tag{3.1}$$

Implicit in our definition is that exercise involves either receiving the net proceeds $(X_\tau - K) > 0$ ("cash" exercise) or discarding the option at expiry as it is worthless (if $X_T < K$). As is standard, we do not allow the agent to exercise before expiry if $X_s < K$. Since $(I_s^{(t)})_{s \geq t}$ is increasing we can write $I_s^{(t)} = \int_t^s dI_u^{(t)}$ with the convention that $I_{t-}^{(t)} = 0$ so that $dI_t^{(t)} = I_t^{(t)}$. This covers the case where an option is exercised at time t, for then $I_t^{(t)} > 0 = I_{t-}^{(t)}$. We also define the constant income process $i^{(t)}$ by $i_s^{(t)}(\omega) = i \ \forall \ s \geq t, \forall \omega$. In particular, $0^{(t)}$ is the zero cashflow.

Let $\mathcal{V}(I^{(t)})$ be the valuation (at time t) by the agent of the random future cashflow $(I_s^{(t)})_{s \geq t}$. Note $\mathcal{V}(I^{(t)})$ is a known quantity at time t. A simple example is the expected value of income received during $[t, T]$, denoted $\mathcal{V}(I^{(t)}) = \mathbb{E}[I_T^{(t)} | \mathcal{F}_t]$. Our assumption is that the agent prefers more to less.

HYPOTHESIS 3.1. *(Prefer more to less) Suppose $(I_s^{(t)})_{s \geq t}$ and $(J_s^{(t)})_{s \geq t}$ are income processes. If $dI_s^{(t)} \geq dJ_s^{(t)}$ for all $\omega \in \Omega$ and for all $s \geq t$ then $\mathcal{V}(I^{(t)}) \geq \mathcal{V}(J^{(t)})$.*

Hypothesis 3.1 is a minimal requirement on preferences which should be valid in all nonpathological circumstances. It is much weaker than pricing by expectation. Later we will impose stronger hypotheses on preferences.

The cashflow resulting from the exercise of a single American call is given in (3.1). We assume that the agent is able to choose the best exercise time and hence the value of the call (with no vesting, or which has already vested) is given by

$$\mathcal{A} = \mathcal{A}(K, T) = \sup_{\tau : t \leq \tau \leq T} \mathcal{V}(I^{(t)}(\tau)), \tag{3.2}$$

where the supremum is taken over stopping times. The value $\mathcal{A}$ depends on current time t, the characteristics of the option K, T, and may in general depend upon the past history of X and other information contained in $\mathcal{F}_t$. Note that since t is fixed throughout, we do not need to (and choose not to) include the parameter t in the notation for $\mathcal{A}$. In particular, we do not think of $\mathcal{A}_t$ as a process indexed by time.

For the American call with vesting, we have

$$\mathcal{A} = \mathcal{A}(K, T, V) = \sup_{\tau : t \vee V \leq \tau \leq T} \mathcal{V}(I^{(t)}(\tau)). \tag{3.3}$$

Note if the option has already vested, i.e., $V \leq t$ then $\mathcal{A}(K, T, V) = \mathcal{A}(K, T)$.

We stress that at this point, we have not made any assumptions beyond Hypothesis 3.1 on the agent's preferences, or any assumptions on the price process itself.

For the situation with no vesting period, the following results are given in Merton (1973) and Cox and Rubinstein (1985). These are classic no-arbitrage results and the (now textbook) arguments to enforce these relations involve the purchase and sale of options, see Cox and Rubinstein (1985, Chapter 4, Propositions 2 and 3) for details. However since we are interested in incomplete market situations where these trades cannot be implemented (such as employee stock options) we need to give alternative proofs which do not rely on trading, but instead rely on the weak preferences in Hypothesis 3.1. These are close in spirit to the arguments of Merton (1973) which rely on the notion of portfolio dominance.

PROPOSITION 3.2.

(i) *Fix K, V. Then $\mathcal{A}(K, T, V)$ is increasing in maturity T.*
(ii) *Fix K, T. Then $\mathcal{A}(K, T, V)$ is decreasing in vesting date V. Suppose $\mathcal{V}$ satisfies Hypothesis 3.1. Then*
(iii) *Fix T, V. $\mathcal{A}(K, T, V)$ is decreasing in K.*

Proof. (i) and (ii). These are immediate from the fact that as T increases or V decreases, the set of admissible strategies increases, so that if $T \le T'$, $\{\tau\colon t\vee V \le \tau \le T\} \subseteq \{\tau\colon t\vee V \le \tau \le T'\}$ and if $t \le V' \le V$, $\{\tau\colon t\vee V \le \tau \le T\} \subseteq \{\tau\colon t\vee V' \le \tau \le T'\}$.

(iii) Let $K' > K$. For any $t\vee V \le \tau \le T$ let $I_s^{(t)}(\tau, K) = (X_\tau - K)^+ 1_{\{s \ge \tau\}}$ and $dI_s^{(t)}(\tau, K) = (X_\tau - K) 1_{\{t\vee V \le \tau \le T\}} 1_{\{\tau = s\}}$. It follows from Hypothesis 3.1 that for any such τ, $\mathcal{V}(I^{(t)}(\tau, K')) \le \mathcal{V}(I^{(t)}(\tau, K))$ and optimizing over τ we conclude $\mathcal{A}(K', T, V) \le \mathcal{A}(K, T, V)$. $\square$

Now we want to fix option characteristics (K, T, V) and vary the current data to consider the optimal exercise decision. Define the exercise set $\mathcal{E}^{(t)} \subseteq \Omega$ at current time t to be[2]

$$\mathcal{E}^{(t)} = \mathcal{E}^{(t)}_{K,T,V} = \begin{cases} \emptyset & t < V \\ \{X_t > K\} \cap \{\mathcal{A}(K, T, V) = \mathcal{A}(K, t)\} & V \le t < T \\ \Omega & t = T \end{cases} \tag{3.4}$$

and the continuation set

$$\mathcal{C}^{(t)} = \mathcal{C}^{(t)}_{K,T,V} = \begin{cases} \Omega & t < V \\ \{X_t \le K\} \cup \{\mathcal{A}(K, T, V) > \mathcal{A}(K, t)\} & V \le t < T. \\ \emptyset & t = T \end{cases} \tag{3.5}$$

We define also the exercise and continuation sets for an American call which has either already vested, or had no vesting period. Define for the current time t

$$\mathcal{E}^{(t)} = \mathcal{E}^{(t)}_{K,T} = \{X_t > K\} \cap \{\mathcal{A}(K, T) = \mathcal{A}(K, t)\}; \quad t < T$$

with $\mathcal{E}^{(T)}_{K,T} = \Omega$ and

[2]Note by convention we require the option to be in-the-money to be in the exercise set. Under the Black–Scholes model this does not need comment, but if the price process is more general, and if the price X can hit zero in finite time (and then zero is absorbing for X) then the option is worthless.

$$\mathcal{C}^{(t)} = \mathcal{C}^{(t)}_{K,T} = \{X_t \le K\} \cup \{\mathcal{A}(K, T) > \mathcal{A}(K, t)\}; \quad t < T$$

with $\mathcal{C}^{(T)}_{K,T} = \emptyset$.

Given that on $(V \le t)$ we have $\mathcal{A}(K, T, V) = \mathcal{A}(K, T)$, it follows immediately that if $t \ge V$

$$\mathcal{C}^{(t)}_{K,T,V} = \mathcal{C}^{(t)}_{K,T}. \tag{3.6}$$

We now define some further useful concepts: a notion of monetary valuation and a stronger hypothesis on preferences.

DEFINITION 3.3. *We say $\mathcal{V}$ is monetary if $\mathcal{V}(0^{(t)}) = 0$ and $I^{(t)}_s = i + J^{(t)}_s$ implies $\mathcal{V}(I^{(t)}) = i + \mathcal{V}(J^{(t)})$, i.e., a cashflow of $i \ge 0$ at time t is given value i.*

Risk neutral pricing and utility indifference pricing with negative exponential utility are both examples of monetary pricing rules. However in general, other utility functions lead to nonmonetary pricing rules—the value to an agent of an income stream can depend on her wealth. Note our use of the terminology monetary is identical to its use in the axiomatic treatment of risk measures, see Föllmer and Schied (2004).[3]

DEFINITION 3.4. *$I^{(t)} \succeq J^{(t)}$ if $I^{(t)}_s(\omega) \ge J^{(t)}_s(\omega)$ for all $s \ge t$ and for all $\omega \in \Omega$.*

HYPOTHESIS 3.5. *(Prefer more to less and prefer cash earlier) Suppose $I^{(t)} \succeq J^{(t)}$. Then $\mathcal{V}(I^{(t)}) \ge \mathcal{V}(J^{(t)})$.*

Hypothesis 3.5 says agents prefer more to less and prefer cash at earlier times. This is consistent with nonnegative interest rates. Clearly this is still a very weak requirement on preferences, but if $\mathcal{V}$ satisfies Hypothesis 3.5 then it automatically satisfies Hypothesis 3.1.

Prior to the main result of this section, we give a simple lemma which is needed later.

LEMMA 3.6. *Suppose $\mathcal{V}$ is monetary. Suppose $t \ge V$. Then $\mathcal{A}(K, T, V) = \mathcal{A}(K, T) \ge (x - K)^+$.*

Proof. Suppose $X_t = x$. The strategy of immediate exercise yields $\mathcal{V}(I^{(t)}(t, K)) = \mathcal{V}(((x - K)^+)^{(t)}) = (x - K)^+$, where we use the monetary property of $\mathcal{V}$. Hence $\mathcal{A}(K, T) \ge \mathcal{V}(I^{(t)}(t, K)) = (x - K)^+$. □

Here is our first major result which says that the continuation regions are nested for different strikes/maturities. Note that this does not prevent them from being complicated in shape, see the counterexamples below and Figures 4.1 and 4.2.

THEOREM 3.7. *Suppose $\mathcal{V}$ is monetary and satisfies Hypothesis 3.5. If $K \le K'$, $T \le T'$ and $V \le V'$ then $\mathcal{C}^{(t)}_{K,T,V} \subseteq \mathcal{C}^{(t)}_{K',T',V'}$*

Proof. Suppose $X_t = x$ and for now, ignore vesting. We prove if $T \le T'$ then $\mathcal{C}^{(t)}_{K,T} \subseteq \mathcal{C}^{(t)}_{K,T'}$. Suppose $\omega \in \mathcal{C}^{(t)}_{K,T}$. Then $t < T \le T'$. Suppose $x > K$. (Otherwise $\omega \in \mathcal{C}^{(t)}_{K,T'}$ by definition.) Then

[3] See also the recent body of work on optimal stopping under dynamic convex risk measures and nonlinear expectations by Bayraktar and Yao (2011), Bayraktar, Karatzas, and Yao (2010), Krätschmer and Schoenmakers (2010), and Riedel (2009).

$$\mathcal{A}(K, t) < \mathcal{A}(K, T) \leq \mathcal{A}(K, T')$$

by Theorem 3.2 (i). Hence $\omega \in \mathcal{C}^{(t)}_{K,T'}$.

Now we prove if $K \leq K'$, then $\mathcal{C}^{(t)}_{K,T} \subseteq \mathcal{C}^{(t)}_{K',T}$. Suppose $\omega \in \mathcal{C}^{(t)}_{K,T}$. Then $t < T$. If $x \leq K'$ then $\omega \in \mathcal{C}^{(t)}_{K',T}$ as the option is out-of-the-money and there is nothing to prove. So suppose $t < T$ and $X_t = x > K'$. There exists τ such that $\mathcal{V}(I^{(t)}(\tau, K)) > \mathcal{V}(I^{(t)}(t, K)) = \mathcal{V}(((x - K)^+)^{(t)}) = (x - K)^+ = x - K$, where we use the monetary property of $\mathcal{V}$. For this τ,

$$J^{(t)}_u(\tau) := I^{(t)}_u(\tau, K') + (K' - K) \geq I^{(t)}_u(\tau, K)$$

and hence $J^{(t)}(\tau) \succeq I^{(t)}(\tau, K)$. By Hypothesis 3.5,

$$\mathcal{V}(I^{(t)}(\tau, K') + (K' - K)) \geq \mathcal{V}(I^{(t)}(\tau, K)).$$

But since $\mathcal{V}$ is monetary, $\mathcal{V}(I^{(t)}(\tau, K') + (K' - K)) = \mathcal{V}(I^{(t)}(\tau, K')) + (K' - K)$ and it follows that

$$\mathcal{A}(K', T) + (K' - K) \geq \mathcal{V}(I^{(t)}(\tau, K)) > (x - K).$$

Hence $\mathcal{A}(K', T) > (x - K')$ so $\omega \in \mathcal{C}^{(t)}_{K',T}$.

Now consider extending the proof to include vesting dates. Set $V \leq V'$. If $\omega \in \mathcal{C}^{(t)}_{K,T,V}$ then by (3.6) either $\omega \in \mathcal{C}^{(t)}_{K,T} \subseteq \mathcal{C}^{(t)}_{K',T'}$ or $t \leq V$ and then $t \leq V'$ and $\mathcal{C}^{(t)}_{K',T',V'} = \Omega$. Hence $\omega \in \mathcal{C}^{(t)}_{K',T',V'}$. □

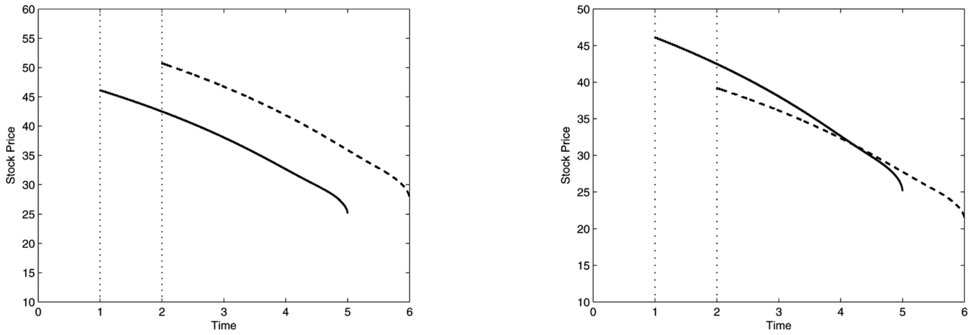
REMARK 3.8. Theorem 3.7 tells us that at the current time t , the continuation set for the option with characteristics K, T, V is contained in that with characteristics K', T', V'. Since t and $\mathcal{F}_t$ are arbitrary, this will remain true at future times s with $t < s \leq T \wedge T'$. In particular if we define the exercise set $\mathcal{E}^{(s)}$ at time s, and if we define (an) optimal stopping time by

$$\tau^* = \tau^*(\omega) = \inf\{s \geq t; \omega \in \mathcal{E}^{(s)}\} \tag{3.7}$$

then under the assumptions of Theorem 3.7, $\tau^* \leq \tau'^*$.

Theorem 3.7 tells us that under valuation methodologies which are monetary and satisfy Hypothesis 3.5, if two identical agents each hold a different American call option, the agent with the lower strike/earlier maturity/shorter vesting date will exercise first. That is, the continuation sets for options with comonotonic strike, expiry, and vesting date are nested. Note this result does not apply to a *portfolio* of options held by a single agent, at least not at this level of generality on preferences. Of course if one were to make the strong assumption of risk neutrality, options can be treated independently, and the result can be applied. Since our main interest is in nonrisk neutral pricing, our aim is to develop a portfolio version of Theorem 3.7 in Section 6.

In the Markov case when $(X_s)_{t \leq s \leq T}$ is a diffusion, it is clear that $\mathcal{A} = \mathcal{A}(x, t)$ and for $t < T$, $\mathcal{E}^{(t)} = \{X_t \in E^{(t)}\}$ for some set $E^{(t)} \subseteq (K, \infty)$ with similar conventions for $\mathcal{C}^{(t)}$. We set $E^{(T)} = [0, \infty)$. In this case it is convenient to consider $\mathcal{A}(X_t, t)$ and $\mathcal{E}^{(t)}$ as t varies, and to think of $\mathcal{A}(x, s)$ as the value function and $E^{(s)} \subseteq \mathbb{R}^+$ as the exercise set at future times



(a) Co-monotonic: $(K = 11, T = 6, V = 2)$ (b) $(K = 8.5, T = 6, V = 2)$

FIGURE 3.1. The exercise threshold of a single American call under risk neutral pricing and exponential Brownian motion. In each panel, the solid line gives the threshold for call with ($K = 10$, $T = 5$, $V = 1$). The dashed line in each panel is the threshold for the call with ($T = 6$, $V = 2$) and different strike in each panel as indicated. Other parameter values are: $r = 0.05$, $q = 0.02$, $\sigma = 0.4$.

s. Then we can define the *exercise region*

$$E = \bigcup_{s \geq t} E^{(s)} \tag{3.8}$$

as a region in $\mathbb{R}^+ \times [t, T]$. We will reserve the terminology region for the Markov case and subsets of $\mathbb{R}^+ \times [t, T]$.

3.1. Example—Black–Scholes Model: Nested and Nonnested Continuation Regions

A first simple example is the Black–Scholes framework where $\mathcal{V}$ is risk neutral pricing and X follows exponential Brownian motion with constant interest rate r, constant proportional dividends q, and constant volatility σ. Figure 3.1 gives exercise thresholds for pairs of calls. We describe how the thresholds are computed in the Appendix, Section (i). The exercise region takes the form

$$E = \{(x, s) : (s \geq t \vee V) \quad \text{and} \quad x \geq x^*(s)\} \cup \{(x, s) : s = T\}$$

for some function of time $x^*(s)$ (see the later result in Theorem 4.2). In each panel, the solid line is the exercise threshold $x^*(s)$ for a call with ($K = 10$, $T = 5$, $V = 1$). Note if X is above the threshold level at the vesting date of 1 year, the call is exercised at the vest date. In panel (a), the dashed line is the threshold for a call with ($K = 11$, $T = 6$, $V = 2$). The continuation regions are nested, as expected, since the strikes, maturities, and vesting dates are comonotonic.

If, for instance, we consider the pair of options with ($K = 10$, $T = 5$, $V = 1$) and ($K = 8.5$, $T = 6$, $V = 2$) then we violate the conditions of the theorem. As we see in panel (ii) of the figure, this results in continuation regions which are not nested.

The proof of Theorem 3.7 relied on the monetary property and Hypothesis 3.5. (An alternative proof could use convexity of $\mathcal{A}$ in strike K, see Hobson 1998, however this relies on expectation pricing, which is a stronger requirement than the monetary property.) The well-known textbook arguments of Cox and Rubinstein (1984; Chapter 4, Proposition 5(c)) use no-arbitrage arguments which involve trading in order to enforce no-arbitrage. The monetary property is, of course, implicit but unstated in such arguments as the potential arbitrage requires the agent to assign a monetary value to a cash profit. We emphasize the importance of the monetary property by the following counterexample.

3.2. Counterexample—Nonmonetary and Nonnested Continuation Regions

Suppose $t = 0$ and $T = 1$. Let $I_s^{(0)}(\tau) = (X_\tau - K)^+ 1_{s \geq \tau}$ be the income from option exercise at $0 \leq \tau \leq T$ and suppose $\mathcal{V}(I^{(0)}) = \mathbb{E}\, U[(X_\tau - K)^+ + w]$ where w is initial wealth. We make a particular choice of U to enable calculations to be carried out. (Note, here for simplicity we define $\mathcal{V}$ as the value function rather than utility indifference price as used later in the paper.) Define $\lfloor x \rfloor$ the integer part of x, and $\mathrm{frac}(x) = x - \lfloor x \rfloor$ the fractional part of x. Let $U(y) = \lfloor y \rfloor$ and suppose $X_s = x$ for $0 \leq s \leq 1$ and $X_1 = x + Z$ where Z takes values $\{\frac{1}{2}, -\frac{1}{2}\}$ with equal probability.

Take $w = 0$ for simplicity and consider $K = 0$. Then it is worth the agent waiting until time 1 to exercise if $\mathrm{frac}(x) \geq 1/2$, otherwise waiting can only reduce the utility. Hence

$$\mathcal{V}(I^{(0)}) = \begin{cases} \lfloor x \rfloor + 1/2 & \mathrm{frac}(x) \geq 1/2 \\ \lfloor x \rfloor & \mathrm{frac}(x) < 1/2 \end{cases}$$

and $\mathcal{C}^{(0)}_{K=0} = \{x : \mathrm{frac}(x) \geq 1/2\}$. Now consider $K = 1/2$. Then

$$\mathcal{V}(I^{(t)}) = \begin{cases} \lfloor x - 1/2 \rfloor & \mathrm{frac}(x) \geq 1/2 \\ \lfloor x \rfloor - 1/2 & \mathrm{frac}(x) < 1/2 \end{cases}$$

and $\mathcal{C}^{(0)}_{K=1/2} = \{x : \mathrm{frac}(x) < 1/2\}$. Then $\mathcal{C}^{(0)}_{K=0}$ and $\mathcal{C}^{(0)}_{K=1/2}$ are disjoint.

This extreme example has been carefully constructed to facilitate computations. However, the intuition remains true with examples with more realistic features. Take U to be an increasing, concave, piecewise linear function. Suppose there we are close to the option maturity (small time to go) and that it is reasonable to consider X to be a Brownian motion with a small positive drift. If the values of strike and wealth put us at or near a kink, then the downside risk dominates and we prefer to stop. If on the other hand, parameters mean we are in a linear part of the function, away from kinks, the positive drift means that we would continue. We can now perturb the strike K in such a way that the stopping and continuation sets are reversed, and thus are not nested.

Our final result in this section tells us when we can expect continuation regions to be nested for different dividend rates. Intuition from Black–Scholes informs us that a higher level of dividends reduces the American call boundary and hence the continuation region. In contrast to Theorem 3.7, we will need some assumptions on the price X. We make:

ASSUMPTION 3.9.

(i) The price X follows a diffusion model:

$$dX_s = X_s(r - q(X_s, s))\, ds + X_s \sigma(X_s, s)\, dB_s, \tag{3.9}$$

where $r > 0$ is a constant interest rate, $q(X_s, s)$ is a nonnegative dividend rate, $\sigma(X_s, s)$ is the volatility and r, q, σ are such that zero is an absorbing point, and that (3.9) is unique in law.

(ii) *Define $M = (M_s)_{s \geq t}$ by*

$$M_s = X_s \, exp\left(-\int_t^s (r - q(X_u, u))\,du\right). \tag{3.10}$$

Then $(M_s)_{s \geq t}$ is a true martingale.

Note, as is common in financial applications, we are interested in weak solutions to (3.9). Further, note that in (3.10) $dM_s = \sigma(X_s, s)M_s dB_s$ so that M is automatically a local martingale. Various tests (e.g., Novikov) exist to ensure M is a martingale. A simple sufficient condition is that $\sigma(X_s, s)$ is bounded.

DEFINITION 3.10. *$\mathcal{V}$ is distribution invariant if whenever the laws of $I^{(t)}$ and $J^{(t)}$ are identical we have $\mathcal{V}(I^{(t)}) = \mathcal{V}(J^{(t)})$.*

Note $I^{(t)}$ and $J^{(t)}$ may be defined on different probability spaces.

THEOREM 3.11. *Suppose Hypothesis 3.1 holds and suppose $\mathcal{V}$ is distribution invariant. Suppose $\hat{X}$ solves*

$$X_s = x + \int_t^s X_u(r - \hat{q}(X_u, s))du + \int_t^s X_u\sigma(X_u, u)d\hat{B}_u \tag{3.11}$$

subject to $X_t = x$, and that Assumption 3.9(i) holds for $\hat{X}$. Suppose $\tilde{X}$ is a solution to

$$X_s = x + \int_t^s X_u(r - \tilde{Q}_u)du + \int_t^s X_u\sigma(X_u, u)d\tilde{B}_u \tag{3.12}$$

with $\tilde{X}_t = x$. Suppose either:

(a) *$x\hat{q}(x, s)$ and $x\sigma(x, s)$ are Lipschitz in the sense that*

$$|x\hat{q}(x, s) - y\hat{q}(y, s)| + |x\sigma(x, s) - y\sigma(y, s)| \leq k|y - x| \tag{3.13}$$

for every $s < \infty$ and $x, y \in \mathbb{R}$ where k is a positive constant, and $\tilde{Q}_s = \tilde{q}(X_s, s)$ for some continuous function $\tilde{q}$ of x and s, which is sufficiently regular that (3.12) has a strong solution; or

(b) *$\sigma(x, s) = \sigma$ (constant volatility), $\hat{q}$ is Lipschitz in log scale in the sense that*

$$|\hat{q}(e^x, s) - \hat{q}(e^y, s)| \leq k|y - x| \tag{3.14}$$

and $\tilde{q}_s$ is such that (3.12) has a weak solution.

Suppose $\tilde{q}_s \geq \hat{q}(\tilde{X}_s, s)$. Then $\tilde{\mathcal{C}}^{(t)} \subseteq \hat{\mathcal{C}}^{(t)}$ where $\tilde{\mathcal{C}}$ and $\hat{\mathcal{C}}$ are the respective continuation regions.

REMARK 3.12.

(i) In case (a) we require both models to be diffusions, although only $\hat{q}$ appears in the Lipschitz condition.

(ii) In case (b) we make no assumption on the drift of $\tilde{X}$ beyond the fact that the dividend rate at time s is bounded below by the Lipschitz function $\hat{q}$. In particular, $\tilde{X}$ need not be a diffusion. Condition (3.14) ensures that (3.11) has a strong solution.

(iii) Both Leung and Sircar (2009) and Carpenter et al. (2010) show higher dividends reduces the continuation region under preferences given by specific utility functions. Both papers assume prices are exponential Brownian motion with proportional dividends so $\sigma(X_s, s) = \sigma$ and $q(X_s, s) = q$.

Proof. First we show that there exists a coupling such that $\hat{X}_s \geq \tilde{X}_s$ almost surely. In case (a) this follows from a standard comparison theorem for SDEs, e.g., Karatzas and Shreve (1994) Proposition 5.2.18. In case (b), let $\tilde{X}$ be a weak solution of (3.12) for the Brownian motion $\tilde{B}$ and let $\bar{X}$ be given by

$$\bar{X}_s = x + \int_t^s \bar{X}_u(r - \hat{q}(\bar{X}_u, s))du + \int_t^s \bar{X}_u \sigma d\tilde{B}_u. \tag{3.15}$$

Then $\bar{X}$ and $\hat{X}$ are equal in law. Let $\bar{Y} = \log \bar{X}$ and similarly for $\hat{Y}$, $\tilde{Y}$. Then

$$\bar{Y}_s = \log x + \int_t^s \left(r - \frac{1}{2}\sigma^2 - \hat{q}(e^{\bar{Y}_u}, u) \right) du + \int_t^s \sigma d\tilde{B}_u$$
$$\tilde{Y}_s = \log x + \int_t^s \left(r - \frac{1}{2}\sigma^2 - \tilde{Q}_u \right) du + \int_t^s \sigma d\tilde{B}_u$$

and we write $Z_s = \tilde{Y}_s - \bar{Y}_s = \int_t^s [\hat{q}(e^{\bar{Y}_u}, u) - \tilde{q}_u]du$. We want to show $Z_s \leq 0$ from which it follows that $\bar{X}_s \geq \tilde{X}_s$. Let $\phi_n(z)$ be given by

$$\phi_n(z) = \begin{cases} 0 & z \leq 0 \\ nz^2/2 & 0 < z \leq 1/n \\ z - 1/2n & z > 1/n \end{cases}.$$

Then $\phi_n(z)$ is such that $0 \leq \phi_n'(z) \leq 1$ and $\phi_n(z) \uparrow z$. We have $Z_t = 0$ and so

$$\begin{aligned}
\phi_n(Z_s) &= \phi_n(Z_t) + \int_t^s \phi_n'(Z_u) dZ_u = \int_t^s \phi_n'(Z_u) 1_{Z_u \geq 0} dZ_u \\
&= \int_t^s \phi_n'(Z_u) 1_{Z_u \geq 0} [\hat{q}(e^{\bar{Y}_u}, u) - \tilde{q}_u] du \\
&= \int_t^s \phi_n'(Z_u) 1_{Z_u \geq 0} [\hat{q}(e^{\bar{Y}_u}, u) - \hat{q}(e^{\tilde{Y}_u}, u) + \hat{q}(e^{\tilde{Y}_u}, u) - \tilde{q}_u] du \\
&\leq k \int_t^s \phi_n'(Z_u) 1_{Z_u \geq 0} |\bar{Y}_u - \tilde{Y}_u| du = k \int_t^s \phi_n'(Z_u) Z_u^+ du \leq k \int_t^s Z_u^+ du.
\end{aligned}$$

Let $n \uparrow \infty$. Then $Z_s^+ \leq k \int_t^s Z_u^+ du$. By Gronwell's lemma, $Z_s^+ = 0$.

For any $V \vee t \leq \tau \leq T$, let $\hat{I}_s^{(t)}(\tau) = (\hat{X}_\tau - K)^+ 1_{s \geq \tau}$ and $\tilde{I}_s^{(t)}(\tau) = (\tilde{X}_\tau - K)^+ 1_{s \geq \tau}$. Then $d\hat{I}_s^{(t)}(\tau) = (\hat{X}_\tau - K) 1_{t \vee V \leq \tau \leq T} 1_{\tau = s}$ and similarly for $d\tilde{I}_s^{(t)}(\tau)$. It follows from Hypothesis 3.1 that for any such τ, $\mathcal{V}(\hat{I}^{(t)}(\tau)) \geq \mathcal{V}(\tilde{I}^{(t)}(\tau))$ and optimizing over τ we conclude $\hat{A}(K, T, V) \geq \tilde{A}(K, T, V)$. Hence if $\omega \in \tilde{\mathcal{C}}^{(t)} = \{\omega : \tilde{\mathcal{A}}(K, T, V) > \tilde{\mathcal{A}}(K, t, V)\}$ then $\omega \in \hat{\mathcal{C}}^{(t)}$. $\square$

4. CHARACTERIZING THE EXERCISE BOUNDARY OF THE AMERICAN CALL

In this section, we will temporarily add more structure to our set-up. This will (i) allow us to give some results on the form of exercise boundaries for single American calls, (ii) demonstrate how complicated the boundaries can be even just for a single option, and (iii) motivate our main result in Section 6—where we will remove the structure and generalize Theorem 3.7 to option portfolios.

For this section we make Assumption 3.9(i) and (ii). As discussed earlier, since $(X_s)_{s\geq t}$ is Markovian, for fixed K, T $\mathcal{A}(K, T, V)$ is a function of current price x and current time t alone, and we write $\mathcal{A}(K, T, V) = \mathcal{A}(x, t; K, T)$ or $\mathcal{A}(x, t)$. Recall the definition of the exercise region $E \subseteq \mathbb{R}^+ \times [t, T]$ from (3.8). Let the continuation region C be given by

$$C = (\mathbb{R}^+ \times [t, T]) \setminus E.$$

Unless otherwise stated, for this section, we assume that value $\mathcal{V}$ is given by the discounted expectation under a martingale measure, and in particular, the risk-neutral expectation. We also assume that the riskless bond pays a positive constant interest rate r, although the results generalize to deterministic interest rates. Then the value of income stream $I^{(t)}$ at current time t is given by

$$\mathcal{V}(I^{(t)}) = \mathbb{E}\left[\int_{s\geq t} e^{-r(s-t)} dI_s^{(t)}\right]$$

and

$$\mathcal{A}(x, t; K, T, V) = \sup_{\tau: V\vee t\leq\tau\leq T} \mathcal{V}(I^{(t)}(\tau)) = \sup_{V\vee t\leq\tau\leq T} \mathbb{E}^{x,t}[e^{-r(\tau-t)}(X_\tau - K)^+], \tag{4.1}$$

where $\mathbb{E}^{x,t}$ denotes dependence on the current time and current price, and where stopping times τ are defined relative to the filtration generated by $(X_s)_{s\geq t}$. Note we are not claiming that this is a good assumption for employee stock options (recall employees cannot hedge perfectly and thus face an incomplete market), but rather we will develop some results first in a risk neutral setting to demonstrate the complexities involved in characterizing boundaries and thus exercise strategies even in this simple setting.

American option pricing (in complete markets) dates back to McKean (1965), Merton (1973), and Van Moerbeke (1976) who recognize the value of an American call on a stock paying continuous dividends as a free boundary problem. Hedging arguments were given by Bensoussan (1984) and Karatzas (1988). Myneni (1992) surveys the development of American option pricing.

We are first interested in when the optimal exercise and continuation regions can be separated by a single price level at each point in time.

DEFINITION 4.1. *We say the exercise region is of single threshold form if for each $s \in (t, T)$, $E^{(s)} = \{x \geq x^*(s)\}$ for some function of time $x^*(s) > K$. Then*

$$E = \{(x, s) : (s \geq t \vee V) \ \ and \ \ x \geq x^*(s)\} \cup \{(x, s) : s = T\}$$

THEOREM 4.2. *Suppose the price X satisfies Assumption 3.9 and that $xq(x, s)$ is increasing in x for each s. Suppose the American call is priced according to (4.1). Then the optimal exercise region is of single threshold form.*

Proof. Suppose $(x, s) \in E$, $x > K$ and $s \geq t \vee V$. We want to show that for $y > x$, $(y, s) \in E$. We write $X_u^{x,s}$ to denote that we start from x at time s. We show we can couple $(X_u^{x,s})_{u \geq s}$ and $(X_u^{y,s})_{u \geq s}$ such that $X_u^{y,s}(\omega) \geq X_u^{x,s}(\omega)$ for all $u \geq s$ and all ω. We take two independent realizations of X, denoted $(\tilde{X}_u^{x,s})_{u \geq s}$ and $(\tilde{X}_u^{y,s})_{u \geq s}$ and define $\tilde{\tau} = \inf_v \{\tilde{X}_v^{y,s} \leq \tilde{X}_v^{x,s}\}$. Define

$$X_u^{x,s} = \begin{cases} \tilde{X}_u^{x,s} & s \leq u \leq \tilde{\tau} \wedge T \\ \tilde{X}_u^{y,s}; & u \geq \tilde{\tau} \wedge T \end{cases}$$

and $X_u^{y,s} = \tilde{X}_u^{y,s}$; $\forall u$. Then $X^{x,s}$ is equal in law to $\tilde{X}^{x,s}$ and $X_u^{y,s}(\omega) \geq X_u^{x,s}(\omega)$ for all $u \geq s$ and all ω. This is a classic example of a Doeblin coupling (see the Introduction and p. 24 of Lindvall 1992 and section V.54 of Rogers and Williams 2000).

For any $\tau \geq s$

$$e^{-r\tau}\left(X_\tau^{y,s} - K\right)^+ \leq e^{-r\tau}\left(X_\tau^{x,s} - K\right)^+ + e^{-r\tau}\left(X_\tau^{y,s} - X_\tau^{x,s}\right).$$

Let $Z_u = e^{-r(u-s)}(X_u^{y,s} - X_u^{x,s})$. Then we write $Z_u = N_u - A_u$ where

$$N_u = \int_s^u \left(X_v^{y,s}\sigma\left(X_v^{y,s}, v\right) - X_v^{x,s}\sigma\left(X_v^{x,s}, v\right)\right)e^{-r(v-s)}dB_v + (y - x)$$

and

$$A_u = \int_s^u e^{-r(v-s)}\left(X_v^{y,s}q\left(X_v^{y,s}, v\right) - X_v^{x,s}q\left(X_v^{x,s}, v\right)\right)dv.$$

For the coupled processes $X_v^{y,s} \geq X_v^{x,s}$ and then by the monotonicity of $xq(x, v)$ we have that A is increasing. Then the local martingale N satisfies $N_u \geq Z_u \geq 0$ so is a supermartingale. Hence Z is a supermartingale and

$$\sup_{\tau \geq s} \mathbb{E} Z_\tau = Z_s = y - x.$$

Then since $(x, s) \in E$, we have

$$\begin{aligned} \sup_{s \leq \tau \leq T} \mathbb{E}e^{-r(\tau-s)}\left(X_\tau^{y,s} - K\right)^+ &\leq \sup_{s \leq \tau \leq T} \mathbb{E}e^{-r(\tau-s)}\left(X_\tau^{x,s} - K\right)^+ \\ &\quad + \sup_{s \leq \tau \leq T} \mathbb{E}e^{-r(\tau-s)}\left(X_\tau^{y,s} - X_\tau^{x,s}\right) \\ &= (x - K) + (y - x) = (y - K). \end{aligned}$$

□

Jacka (1991) (under exponential Brownian motion) and Babilua et al. (2007) (under time homogeneous diffusions) proved similar results for American puts without dividends. Göttsche and Vellekoop (2011) study the impact of discrete dividends on the boundary in an exponential Brownian motion model. Finally, Vellekoop and Nieuwenhuis (2009) prove results in a semimartingale model under certain conditions (which play the role of our condition on $q(x, t)$ and the no-crossing property for diffusions). However their conditions would seem difficult to verify in any model outside the diffusion framework above.

Theorem 4.2 can also be deduced via convexity of the American call option value in x, under the (stronger) assumption of proportional dividends. The convexity property will follow a lemma which says the call cannot be worth more than the current stock price.

LEMMA 4.3. *Suppose X satisfies Assumption 3.9 and the American call is priced according to (4.1). Then $\mathcal{A}(x, t, K, T, V) \le x$.*

Proof.

$$\mathbb{E}^{x,t}\big[e^{-r(\tau-t)}(X_\tau - K)^+\big] \le \mathbb{E}^{x,t}[e^{-r(\tau-t)}X_\tau] \\ \le \mathbb{E}^{x,t}\big[e^{-\int_t^\tau (r-q(X_u,u))du}X_\tau\big] = \mathbb{E}^{x,t}[M_\tau] = x$$

since M is a martingale by Assumption 3.9. Hence $\mathcal{A}(x, t, K, T, V) = \sup_{V\vee t\le\tau\le T} \mathbb{E}^{x,t}[e^{-r(\tau-t)}(X_\tau - K)^+] \le x$. □

When (4.1) does not hold, an easy counterexample to the lemma is to consider nonrisk-neutral expectation pricing where X has a large positive drift, e.g., pricing by expectation under the real world measure. Then M is a submartingale and the option is worth more than the current stock price.

PROPOSITION 4.4. *Suppose the price X satisfies Assumption 3.9 and dividends are proportional so*

$$dX_s = (r-q)X_s ds + \sigma(X_s, s)X_s dB_s.$$

Suppose the American call is priced according to (4.1). Then $\mathcal{A}(x, t; K, T, V)$ is convex in x.

Proof. Recall if $M_s^{x,t} = e^{-(r-q)(s-t)}X_s^{x,t}$ then $dM_s^{x,t} = M_s^{x,t}\sigma(X_s^{x,t}, s)dB_s = M_s^{x,t}\tilde{\sigma}(M_s^{x,t}, s)dB_s$ (where $\tilde{\sigma}(m, s) = \sigma(e^{(r-q)(s-t)}m, s)$) and by Assumption 3.9, $M^{x,t}$ is a martingale.

Define $w(m, u) = (e^{-q(u-t)}m - e^{-r(u-t)}K)^+$. Then w is convex in m and $w(M_u^{x,t}, u) = e^{-r(u-t)}(X_u^{x,t} - K)^+$. For $s \ge t$ define

$$v(m, s) = \sup_{\tau\ge s}\mathbb{E}\big[w\big(M_\tau^{m,s}, \tau\big)\big].$$

Then v is a supermartingale and a martingale for $s \le \tau^*$ the optimal strategy.

We again use a Doeblin coupling of diffusions. Let $0 < z < y < u$ and for independent Brownian motions α, β, γ define processes $M^{u,t}, M^{y,t}, M^{z,t}$ via

$$dM_s^{u,t} = \tilde{\sigma}\big(M_s^{u,t}, s\big)d\alpha_s; \quad M_t^{u,t} = u$$
$$dM_s^{y,t} = \tilde{\sigma}\big(M_s^{y,t}, s\big)d\beta_s; \quad M_t^{y,t} = y$$
$$dM_s^{z,t} = \tilde{\sigma}\big(M_s^{z,t}, s\big)d\gamma_s; \quad M_t^{z,t} = z.$$

Let τ^y be the optimal exercise time for the option with price $M^{y,t}$ started at y. Then $V \le \tau^y \le T$.

Define $H_u = \inf\{s : M_s^{u,t} = M_s^{y,t}\}$, $H_z = \inf\{s : M_s^{z,t} = M_s^{y,t}\}$ and $\tau = H_u \wedge H_z \wedge \tau^y$.

On $\tau = \tau^y$, $v(M_\tau^{y,t}, \tau) = w(M_\tau^{y,t}, \tau) = (e^{-q(\tau-t)}M_\tau^{y,t} - Ke^{-r(\tau-t)})$ and by convexity,

$$\big(M_\tau^{u,t} - M_\tau^{z,t}\big)v\big(M_\tau^{y,t}, \tau\big) \le \big(M_\tau^{y,t} - M_\tau^{z,t}\big)w\big(M_\tau^{u,t}, \tau\big) + \big(M_\tau^{u,t} - M_\tau^{y,t}\big)w\big(M_\tau^{z,t}, \tau\big) \\ \le \big(M_\tau^{y,t} - M_\tau^{z,t}\big)v\big(M_\tau^{u,t}, \tau\big) + \big(M_\tau^{u,t} - M_\tau^{y,t}\big)v\big(M_\tau^{z,t}, \tau\big).$$

On $\tau = H_u$, by symmetry

$$\left(M_\tau^{u,t} - M_\tau^{z,t}\right)v\left(M_\tau^{y,t}, \tau\right) = \left(M_\tau^{y,t} - M_\tau^{z,t}\right)v\left(M_\tau^{u,t}, \tau\right)$$

and $(M_\tau^{u,t} - M_\tau^{y,t})v(M_\tau^{z,t}, \tau) = 0$. Similarly, on $\tau = H_z$,

$$\left(M_\tau^{u,t} - M_\tau^{z,t}\right)v\left(M_\tau^{y,t}, \tau\right) = \left(M_\tau^{u,t} - M_\tau^{y,t}\right)v\left(M_\tau^{z,t}, \tau\right).$$

Hence we always have

$$\left(M_\tau^{u,t} - M_\tau^{z,t}\right)v\left(M_\tau^{y,t}, \tau\right) \leq \left(M_\tau^{y,t} - M_\tau^{z,t}\right)v\left(M_\tau^{u,t}, \tau\right) + \left(M_\tau^{u,t} - M_\tau^{y,t}\right)v\left(M_\tau^{z,t}, \tau\right).$$

Taking expectations and by independence,

$$(u - z)\mathbb{E}v\left(M_\tau^{y,t}, \tau\right) \leq (y - z)\mathbb{E}v\left(M_\tau^{u,t}, \tau\right) + (u - y)\mathbb{E}v\left(M_\tau^{z,t}, \tau\right).$$

Now $v(M_s^y, s)$ is a martingale on $t \leq s \leq \tau \leq \tau^y$ so $\mathbb{E}v(M_\tau^{y,t}, \tau) = v(y, t)$; and $v(M_s^{u,t}, s)$, $v(M_s^{z,t}, s)$ are supermartingales, so the convexity property for $v(., t)$ follows. □

This proof generalizes option price convexity results by Bergman, Grundy, and Wiener (1996) for European options (using analysis of the partial differential equation), and El Karoui, Jeanblanc-Picque, and Shreve (1998) (using stochastic flows) and Hobson (1998) (using coupling) (also Ekström 2004) for American options without dividends. In particular, our proof extends the coupling proof of Hobson (1998).

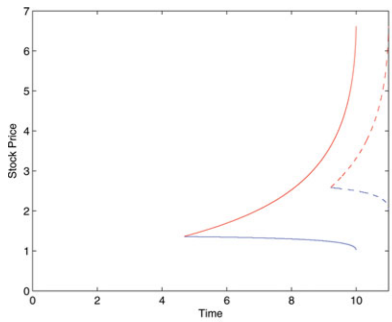
Alternative proof of Theorem 4.2: Fix K, T, V, t. From Proposition 4.4, $\mathcal{A}(x, t)$ is convex in x. Since from Lemmas 3.6 and 4.3, $(x - K)^+ \leq \mathcal{A}(x, t) \leq x$, then if for any $x > 0$ we have $\mathcal{A}(x, t) = (x - K)$, then, for all $y > x$ $\mathcal{A}(y, t) = (y - K)$. Thus if $(x, t) \in E_{K,T}$ then $(y, t) \in E_{K,T}$. □

4.1. Counterexample: Disconnected Exercise Regions

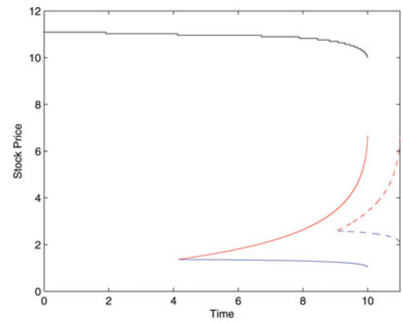
We adopt the Black–Scholes framework. Suppose X follows exponential Brownian motion with constant proportional dividends and suppose pricing is risk-neutral as in (4.1). Under this standard model, the American call has a well-studied exercise boundary, which by our definition is a single threshold (see Figure 3.1). Now we show that if $xq(x, s)$ is not monotonic increasing, then the exercise boundary need not be a single threshold. Suppose that dividends are state-dependent and

$$q = \begin{cases} q_{high}; & X_s \leq x_q \\ q_{low}; & X_s > x_q \end{cases}$$

for some constant x_q. This violates the assumption in Theorem 4.2. Figure 4.1 takes $q_{high} = 0.1$, $x_q = 7$ and assumes $V = 0$. The boundaries are computed using the algorithm described in Section (i) of the Appendix, where the constant proportional dividend q takes the two values as described, depending on which part of the grid of prices X we are in. In panel (a), we take $q_{low} = 0$, whilst in panel (b), $q_{low} = 0.01$. In each panel, the solid lines delineate the exercise and continuation regions for a call with $K = 1$, $T = 10$. We first describe the graph in panel (a). Far from maturity, it is optimal to continue at all price levels. There is some prospect of prices reaching the zero dividend region, and zero



(a) Exercise region if $q_{low} = 0$



(b) Exercise region if $q_{low} = 0.01$

FIGURE 4.1. The figures show the optimal exercise and continuation regions of a single option in a Black–Scholes model with state-dependent dividends. We take $x_q = 7$, $q_{high} = 0.1$. In both panels, the solid line encloses an exercise region for a call with $K = 1$, $T = 10$. In panel (b), there is also a disconnected exercise region at high price levels. In both panels, the dashed line encloses the exercise region for a call with $K = 2$, $T = 11$. Other parameters are: $r = 0.1$, $\sigma = 0.15$, $V = 0$.

dividend American calls are not exercised early (Merton 1973). As maturity approaches, the chance of reaching the zero-dividend region becomes less likely, and there is a wedge of price levels between which it is optimal to exercise. This exercise wedge becomes larger as maturity approaches, and close to maturity it is optimal to exercise for all prices in-the-money. The presence of dividends (for $X > x_q$) in panel (b) induces a disconnected exercise region at high prices, and the option is exercised above the higher solid line.

The dashed lines on Figure 4.1 display the exercise and continuation regions for a call with ($K = 2$, $T = 11$). The exercise region for this call is nested within the region for the call with lower strike and maturity, which is consistent with Theorem 3.2. An agent with the ($K = 1$, $T = 10$) option will exercise before an identical agent with the ($K = 2$, $T = 11$) option. In addition, since this example assumes risk-neutral pricing, we can also conclude that an agent with a portfolio of these two options would exercise the lower strike/maturity option first.

Our next task is to recall some known results concerning the monotonicity of the boundary in time. We need some preliminaries.

PROPOSITION 4.5. *Assume the price X satisfies Assumption 3.9 and in addition, that the dynamics are time-homogeneous, i.e.,* $q(x, s) = q(x)$, $\sigma(x, s) = \sigma(x)$. *Assume the American call is priced according to (4.1). Fix K,T,x. Then*

(i) $\mathcal{A}(x, t; K, T) = \mathcal{A}(x, 0; K, T - t)$;
(ii) For $t' > t$, $\mathcal{A}(x, t; K, T) \geq \mathcal{A}(x, t'; K, T)$

Recall Theorem 3.2(i) says the option value is increasing in maturity whereas Proposition 4.5 (ii) requires time-homogeneity. In contrast to Proposition 4.5 (ii), the European call option value is not nonincreasing with t.

Proof.

(i) Let $\tilde{X}_u = X_{u+t}$ for $0 \leq u \leq T - t$. The filtration $(\mathcal{F}_s)_{t \leq s \leq T}$ where $(\mathcal{F}_s) = \sigma((X_v)_{t \leq v \leq s})$ can be identified with $(\tilde{\mathcal{F}}_s)_{0 \leq s \leq T-t}$ where $(\tilde{\mathcal{F}}_{s-u}) = \sigma((\tilde{X}_v)_{0 \leq v \leq s-u})$. Let $\tilde{\tau} = \tau - t$.

Then

$$e^{-r(\tau-t)}(X_\tau - K)^+ = e^{-r\tilde{\tau}}(\tilde{X}_{\tilde{\tau}} - K)^+.$$

Since the dynamics of X are time-homogeneous, $\tilde{X}$ solves the same SDE as X, and

$$\begin{aligned}\sup_{t\le\tau\le T} \mathbb{E}^{x,t}[e^{-r(\tau-t)}(X_\tau - K)^+] &= \sup_{0\le\tilde{\tau}\le T-t} \mathbb{E}^{x,0}[e^{-r\tilde{\tau}}(\tilde{X}_{\tilde{\tau}} - K)^+]\\ &= \sup_{0\le\tau\le T-t} \mathbb{E}^{x,0}[e^{-r\tau}(X_\tau - K)^+]\end{aligned}$$

and $\mathcal{A}(x, t; K, T) = \mathcal{A}(x, 0; K, T-t)$.

(ii) Let $t < t'$. From Theorem 3.2 (i), $\mathcal{A}(x, 0; K, T-t) \ge \mathcal{A}(x, 0; K, T-t')$, so that from part (i) of this theorem, $\mathcal{A}(x, t; K, T) \ge \mathcal{A}(x, t'; K, T)$. □

THEOREM 4.6. *Suppose X satisfies Assumption 3.9 and in addition that the dynamics are time homogeneous, i.e.,* $q(X_s, s) = q(X_s)$, $\sigma(X_s, s) = \sigma(X_s)$. *Suppose the American call is priced according to (4.1). Assume* $xq(x)$ *increasing in x so (by Theorem 4.2) the exercise region takes the form:*

$$E_{K,T} = \{(x, s) : x \ge x^*(s)\} \cup \{(x, s) : s = T\}.$$

Then

(i) $x^*(s)$ *is nonincreasing*;

(ii) $x^*(s) = f(T-s)$ *for some function f with* $f(u)$ *nondecreasing in u,* $f(u)$ *does not depend on T, but does depend on K and* $f(0) \ge K$.

Proof.

(i) Fix K, T. Since by Proposition 4.5(ii), $\mathcal{A}$ nonincreasing in s, we have that if $(x, s) \in C_{K,T}$ then $(x, s') \in C_{K,T} \forall s' < s$ since $\mathcal{A}(x, s'; K, T) \ge \mathcal{A}(x, s; K, T) > (x-K)^+$. Define $\tilde{H}(x) = \tilde{H}(x; K, T) = \sup\{u : \mathcal{A}(x, u; K, T) > (x-K)^+\}$. Then $x^* = \tilde{H}^{-1}$ and $\tilde{H}$ is nonincreasing.

(ii) Fix K. Define $H(x) = \inf\{u : \mathcal{A}(x, 0; K, u) > (x-K)^+\}$. Since by Theorem 3.2(i), $\mathcal{A}$ increasing in u, for all $u > H(x)$ we have $\mathcal{A}(x, 0; K, u) > (x-K)^+$. Fix T. Then by Proposition 4.5(i) and time-homogeneity,

$$\begin{aligned}\tilde{H}(x) &= \sup\{u : \mathcal{A}(x, 0; K, T-u) > (x-K)^+\}\\ &= T - \inf\{v : \mathcal{A}(x, 0; K, v) > (x-K)^+\} = T - H(x).\end{aligned}$$

Hence $H(x) = T - \tilde{H}(x)$ is nondecreasing and has an inverse which we write as $f = H^{-1}$. Then $x^*(s) = x \Leftrightarrow \tilde{H}(x) = s \Leftrightarrow H(x) = T - s \Leftrightarrow f(T-s) = x$ and hence $x^*(s) = f(T-s)$. □

The intuition is that since the option value is decreasing in time (for a fixed stock price) and the payoff for immediate exercise is time-independent, the exercise boundary will decrease with time. We include Theorem 4.6 to draw a contrast to the boundaries calculated via utility indifference pricing in the next section, where the above intuition breaks down. Further, despite this result being known for many years, it has not always been stated correctly in the literature. Kim (1990) gives this result without the condition of time-homogeneity. His Proposition 1 mis-applies the result in Theorem 3.2 (i), which,

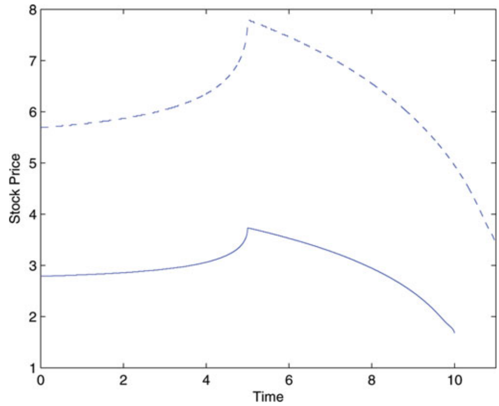


FIGURE 4.2. The figure shows the optimal exercise threshold of a single American call option under risk neutral pricing with a Black–Scholes price with constant proportional dividends. For $s \leq T_\sigma = 5$, $\sigma_{low} = 0.2$, and for $s > T_\sigma = 5$, $\sigma_{high} = 0.4$. The solid line depicts the exercise threshold of an option with $K = 1$, $T = 10$. The dashed line depicts the exercise threshold of an option with $K = 2$, $T = 11$. Other parameters are: $r = 0.05$, $q = 0.03$.

as Cox and Rubinstein (1985) caution, compares the values on a given calendar date of two calls with different maturities. It does not tell us anything about how the value of a call changes with time. Indeed, it is only equivalent to comparing different time-to-expiries with fixed expiry date under the additional assumption of time-homogeneity of the model.

4.2. Counterexample: Nonmonotone Exercise Boundary

Again we take the Black–Scholes setup—risk neutral pricing and exponential Brownian motion with constant proportional dividends. Suppose volatility takes two distinct values $\sigma = \sigma_{low}$, $s \leq T_\sigma$ and $\sigma = \sigma_{high}$, $s > T_\sigma$ for some fixed $T_\sigma \leq T$. Figure 4.2 gives the resulting nonmonotonic exercise boundary for a call with ($K = 1$, $T = 10$) (the solid line). Again, we modify the algorithm in the Appendix, Section (i) to take account of the two values of the volatility. As $s \to T$, the boundary approaches the limit $\frac{r}{q}K = 1.66$. As $s \to -\infty$, the boundary approaches the perpetual limit. The dashed line gives the exercise boundary for a single option with ($K = 2$, $T = 11$). Despite the lack of time homogeneity, and resulting nonmonotonic boundaries, we see the boundaries are consistent with Theorem 3.7 and the continuation regions are nested. Since the example is under risk-neutral pricing, an agent with a portfolio of both options would exercise the ($K = 1$, $T = 10$) option first. Another situation where the exercise boundary is not monotone in time (under risk neutral pricing) is found in Göttsche and Vellekoop (2011) where dividends are paid discretely.

5. SINGLE AMERICAN OPTIONS—UTILITY INDIFFERENCE PRICING

In this section, we consider the exercise of single American options under the utility indifference framework. As we described in Section 2, utility-based models have become a standard framework in which to value employee stock options due to the restrictions on the employee's ability to hedge.

The agent has initial wealth w, increasing concave utility function U, and (for now) a single American call option with strike K, maturity T, vesting date V. In addition to the riskless bond (with positive constant interest rate r) and the underlying stock X (which cannot be traded), there is also a market asset M in which the employee may partially hedge her risk.

Prices follow

$$\frac{dX}{X} = (\nu - q)\,dt + \sigma\,dB \tag{5.1}$$

$$\frac{dM}{M} = \mu^M dt + \sigma^M dZ, \tag{5.2}$$

where standard Brownian motions B and Z with constant instantaneous correlation $\rho \in (-1, 1)$ are defined on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_u\}_{u\geq t}, \mathbb{P})$ and where $\mathcal{F}_u$ is the augmented σ-algebra generated by $\{B_u, Z_u;\ 0 \leq u \leq t\}$. The volatility of stock returns σ, expected return on the stock ν, constant proportional dividend yield $q > 0$, and expected return μ^M and volatility of the market returns σ^M are all constants. The mean stock return ν is equal to the CAPM return for the stock, given its correlation with the market,[4]

$$\nu = r + \beta(\mu^M - r); \quad \beta = \rho\sigma/\sigma^M.$$

Since $\rho \in (-1, 1)$, the restricted employee faces some unhedgeable risk through her option position. Allowing the executive to trade in the market asset enables her to partially hedge this risk. Suppose she holds a cash amount θ_s in M at time s (satisfying the integrability condition $\mathbb{E}\int_t^T \theta_u^2 du < \infty$) and invests the remainder of her wealth at the riskless rate r. The employee's wealth $W = W^{\theta, I^{(t)}}$ consists of two parts—her trading wealth from investment in the market asset and a cashflow $I^{(t)}$ which later will represent the cash income from option exercise. We have

$$dW_u = (rW_u + \theta_u(\mu^M - r))du + \theta_u\sigma^M dZ_u + dI_u^{(t)};\ W_t = w. \tag{5.3}$$

The employee's goal is to maximize expected utility of terminal wealth at some future date T (which we take to be the maturity of the option), and maximization is taken over the choice of adapted hedging strategies $\theta = (\theta_u)_{t\leq u\leq T}$. Define

$$\bar{U}(w) = \bar{U}(w; t) = \sup_\theta \mathbb{E}[U(W_T^{\theta;0}) | W_t = w]$$

which is the value function or indirect utility for the terminal wealth obtainable without any income ($I^{(t)} = 0$). Then the certainty equivalent value of the income $\mathcal{V}(I^{(t)})$ is given

[4]Given the stock is a traded asset, if the CAPM relation did not hold, we would have arbitrage possibilities when $\rho = \pm 1$. See Davis (1999). The CAPM drift choice is natural as it leads to montonicity of the (CARA) utility indifference price in $|\rho|$ (see Henderson 2005, Grasselli and Henderson 2009, and Frei and Schweizer 2008 in a non-Markovian model with stochastic correlation). Leung and Sircar (2009) do not require the CAPM drift to hold, and find an asymmetry in the prices and exercise boundaries for positive and negative correlations of the same magnitude. As argued by Carpenter et al. (2010), "this is because they hold the mean return on the stock fixed as they vary correlation, so that the stock has an abnormal return with respect to the hedging instrument, which is larger the smaller or more negative, the correlation."

by the solution to

$$\bar{U}(w + \mathcal{V}(I^{(t)})) = \sup_{\theta} \mathbb{E}\big[U\big(W_T^{\theta, I^{(t)}}\big) | W_t = w\big]$$

and thus

$$\mathcal{V}(I^{(t)}) = \bar{U}^{-1}\left(\sup_{\theta} \mathbb{E}\big[U\big(W_T^{\theta, I^{(t)}}\big) | W_t = w\big]\right) - w. \tag{5.4}$$

PROPOSITION 5.1. *For $\mathcal{V}$ given in (5.4), Hypothesis 3.5 holds. Moreover $\mathcal{V}$ is distribution invariant.*

Proof. It is clear from (5.3) that $W_T^{\theta, I^{(t)}}$ is increasing in w, so that since U is increasing in w, so is $\bar{U}$. It is sufficient to show that if $I_u^{(t)} \geq J_u^{(t)} \quad \forall u$ then for each θ, $W_T^{\theta, I^{(t)}} \geq W_T^{\theta, J^{(t)}}$, then since U and $\bar{U}$ are increasing, we have $\mathcal{V}(I^{(t)}(\tau)) \geq \mathcal{V}(J^{(t)}(\tau))$ from (5.4) as required. In fact, $W_s^{\theta, I^{(t)}} - W_s^{\theta, J^{(t)}} = \int_t^s r(W_u^{\theta, I^{(t)}} - W_u^{\theta, J^{(t)}})du + I_s^{(t)} - J_s^{(t)}$ so that $W_s^{\theta, I^{(t)}} \geq W_s^{\theta, J^{(t)}}$ for each s. $\mathcal{V}$ is distribution invariant from the fact that valuations are based on expected utility. □

Note $\mathcal{V}$ solving (5.4) is not monetary unless U has CARA.

Now assume $I_s^{(t)}(\tau) = (X_\tau - K)^+ 1_{s \geq \tau}$ is the income from option exercise at $V \vee t \leq \tau \leq T$. Optimizing over the exercise time, the utility indifference value of the option is

$$\mathcal{A} = \mathcal{A}(w, x, t) = \sup_{\tau} \mathcal{V}(I^{(t)}) = \bar{U}^{-1}\left(\sup_{\theta} \sup_{\tau} \mathbb{E}[U\big(W_T^{\theta, I^{(t)}}\big) | W_t = w,\, X_t = x]\right) - w. \tag{5.5}$$

First, we return to the dividend comparison result of Theorem 3.11 and show it holds in the context of utility indifference pricing. This example recovers the result of Carpenter et al. (2010).

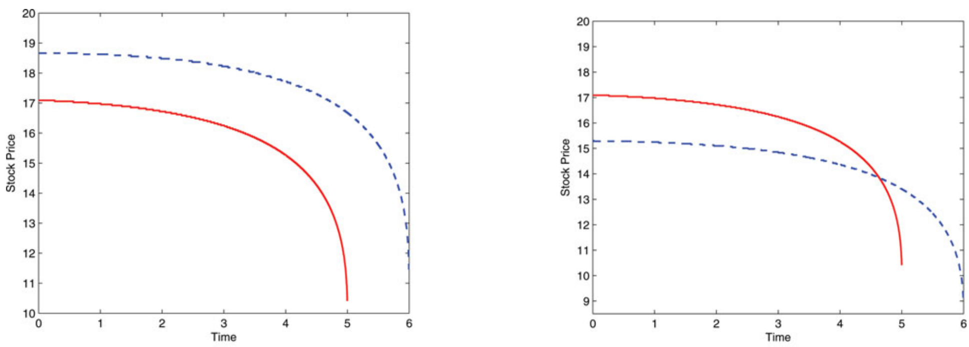
PROPOSITION 5.2. *Suppose $\hat{X}$ solves (5.1) with $\hat{q}$ and $\tilde{X}$ solves (5.1) with $\tilde{q}$, and suppose $\tilde{q} \geq \hat{q}$. For $\mathcal{V}$ given in (5.4), the conclusion of Theorem 3.11 holds: $\tilde{C} \subseteq \hat{C}$.*

Proof. If we define $\hat{X}$ with $\hat{q}$ and $\tilde{X}$ with $\tilde{q}$ as solutions to (5.1) with respect to the same Brownian motion B, then clearly $\hat{X}_s \geq \tilde{X}_s \forall s$. Then $\hat{I}_s^{(t)}(\tau) \geq \tilde{I}_s^{(t)}(\tau)$ as in Theorem 3.11. Then, Proposition 5.1 says Hypothesis 3.5 holds, hence Hypothesis 3.1 holds. We have $\mathcal{V}(\hat{I}^{(t)}(\tau)) \geq \mathcal{V}(\tilde{I}^{(t)}(\tau))$ and thus $\hat{\mathcal{A}}(x, t, K, T, V) \geq \tilde{\mathcal{A}}(x, t, K, T, V)$. Hence if $(x, t) \in \tilde{C}$ then $(x, t) \in \hat{C}$ and $\tilde{C} \subseteq \hat{C}$. □

We now consider two specifications for U and solve the dynamic programming problem resulting from (5.5). Note that since (5.5) is treating a single option, although we display thresholds for different option characteristics, these should be interpreted to be for stand-alone or individual options, rather than portfolios.

5.1. CARA Utility

We specialize to take $U(x) = -e^{-\gamma x}$; $\gamma > 0$ and illustrate the optimal exercise boundaries in Figure 5.1. The free boundary problem resulting from (5.5) is stated in the



(a) Co-monotonic: $(K = 11, T = 6, V = 0)$

(b) $(K = 8.5, T = 6, V = 0)$

FIGURE 5.1. The exercise threshold of a single American call calculated from (5.5) by dynamic programming under $U(x) = -e^{-\gamma x}$. In each panel, the solid line gives the threshold for call with ($K = 10$, $T = 5$, $V = 0$). The dashed line in each panel is the threshold for the call with ($T = 6$, $V = 0$) and different strike in each panel as indicated. Other parameter values are: $\gamma = 0.1$, $\rho = 0.5$, $r = 0.05$, $q = 0.02$, $\sigma = 0.4$.

Appendix (Section (ii)) together with a brief description of the numerical scheme. The Appendix also details the transformation such that the wealth state variable is removed. Leung and Sircar (2009) and Grasselli and Henderson (2009) solve similar free boundary problems. We see the exercise and continuation region are separated by a single threshold, as was the case under the risk-neutral setting in Theorem 4.2. In contrast to Theorem 4.6, the threshold does not have to be monotone in time (see also Rogers and Scheinkman 2007).

Since $\mathcal{V}$ is monetary, we expect the continuation regions for different strikes and maturities to be nested according to Theorem 3.7. Panel (a) shows that this is indeed the case. Of course, if strikes and maturities are not comonotonic then the thresholds may intersect, as shown in Panel (b).

We can compare the exercise thresholds to the equivalent Black–Scholes complete market thresholds in Figure 3.1. As we would anticipate, the thresholds from the utility indifference model are lower than those from the risk neutral model. The agent is not willing to wait for the price to rise as high before exercising because she is risk averse.

Our next example (extended from Carpenter et al. 2010) shows that in contrast to the example with CARA preferences, we need not have a single threshold separating the continuation and exercise regions. Carpenter et al. (2010, proposition 3) give conditions on wealth and the utility function such that a single threshold does indeed separate the continuation and exercise regions. Their conditions on the utility function are easily verified for CARA utility and for CRRA utility with the coefficient of relative risk aversion less than or equal to one. This is an important observation from the perspective of employee stock option modeling, as there is a strand of literature (see, e.g., Cvitanic, Wiener, and Zapatero 2008) which assumes an exogenous form for a boundary of threshold form, rather than deriving the threshold(s) endogenously via a utility (or other) model. It is therefore important to recognize the limitations of such an exogenous specification.

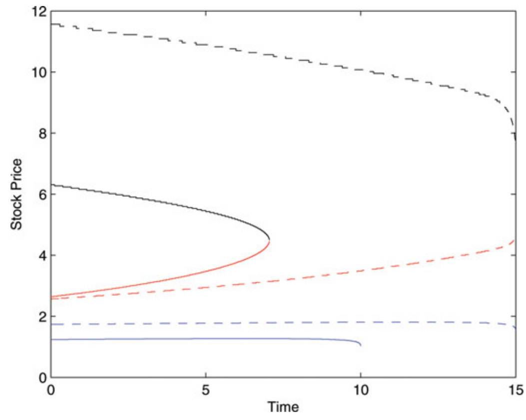


FIGURE 5.2. The optimal exercise region(s) when pricing according to (5.5) with $U(W) = \frac{W^{1-A}}{1-A} + cW$. The solid lines give boundaries for an American call with ($K = 1$, $T = 10$). The dashed lines give boundaries for an American call with ($K = 1.5$, $T = 15$). Note the two options are being treated independently and thus the boundaries do not apply to a portfolio of both options. Other parameters are: $r = 0.05$, $q = 0.01$, $\sigma = 0.6$, $A = 10$, $c = 0.0001$, $w = y = 1.2$.

5.2. Disconnected Continuation Regions

We again solve the dynamic programming problem from (5.5) with $U(W) = \frac{W^{1-A}}{1-A} + cW$. We simplify the calculations by setting $\theta = 0$, i.e., we do not allow hedging in the asset M. This does not alter the observations we wish to make. The details of the resulting free boundary problem are given in Section (ii)(b) of the Appendix.

The resulting thresholds are displayed in Figure 5.2. Consider the dashed lines which give boundaries for a call with $K = 1.5$, $T = 15$. We first describe the graph. Starting from low prices and working upwards to high prices—there is a continuation region for low prices, then an exercise region, then a second continuation region, and then a second exercise region at high prices. The intuition is as follows. For low prices/low wealths, U is like CRRA utility, so exercise occurs when prices become too large relative to wealth and the position is too risky. This explains the presence of the lowest exercise threshold. The continuation region at higher price levels arises from the influence of the risk neutral part of the utility. Dividends are very small (close to zero) and hence a risk neutral agent has an incentive to continue. Indeed, if dividends were zero, this would be the whole story as the agent would continue at high prices everywhere. However, the presence of small dividends induces an exercise region at very high prices. The set of solid lines on the figure give boundaries for an American call option with $K = 1$, $T = 10$. We see that the continuation region for higher prices does not exist during the whole life of the option and is not present as we approach maturity. We stress that the two options have been treated independently and thus the figure does not give any information about a portfolio of the two options. Observe that the conclusions of Theorem 3.7 hold and the continuation regions are nested (i.e., an agent with a low strike/maturity option would exercise before an identical agent with a high strike/maturity option) even though the hypotheses of Theorem 3.7 are not satisfied as the valuation is not monetary. Thus although counterexamples to the nesting of continuation regions exist, such nesting remains typical behavior.

6. PORTFOLIOS OF OPTIONS

In this section, we consider the exercise of portfolios of American call options. We have in mind the situation described earlier, where an employee is granted options periodically and thus builds up an inventory of American call options with varying strikes, vesting dates, and maturities. Of course, if we were content to assume risk neutrality, a portfolio can be considered as a set of independent options each with its own exercise strategy and value. In this case, Theorem 3.7 can be applied. However, we are interested in incomplete market situations whereby perfect hedging is not available. An important consequence is that we cannot decompose a portfolio into a set of independent options—rather the agent's strategy concerning an option will be altered by the existence of other options in her portfolio. This was shown to be the case in Grasselli and Henderson (2009) and Leung and Sircar (2009) where portfolios of options with identical characteristics were considered under CARA preferences. Each option was shown to be exercised at a different threshold level as the agent chooses to unwind risk over time. Note that under risk neutrality, a set of identical options would all be exercised at a single threshold.

We aim to develop an exercise ordering result for portfolios, similar in spirit to the earlier Theorem 3.7, which holds for general preferences, including under utility indifference pricing. The previous section showed exercise and continuation regions from utility indifference pricing may be complex, and that it is difficult to say much in generality. This guides us away from analyzing boundaries and toward finding a model-free approach. *In this section we make no assumptions on the price process X and return to the setup of Section 3*. Then, given a portfolio of American call options with comonotonic strikes, maturities, and vesting dates, we prove that optimal behavior *always* involves exercising the shorter-dated, lower-strike option first. This can be shown using a coupling construction where it is shown that an agent who exercises in order of increasing strike/maturity always generates an amount of wealth that dominates that generated by an agent who follows any other exercise strategy.

Consider an agent with a portfolio of American-style call options. The portfolio consists of n options with characteristics $(K^{(i)}, T^{(i)}, V^{(i)})_{1 \leq i \leq n}$ with vesting dates such that $t \vee V^{(i)} \leq T^{(i)}$. If the option with label i has no vesting, or has already vested, then it is equivalent to suppose that the vesting date $V^{(i)}$ is the current time t. We consider an exercise strategy $\mathcal{S}$ to be a collection of m stopping times (with $0 \leq m \leq n$), and associated labels so that $\mathcal{S} = \{(\tau_i, \ell_i)_{1 \leq i \leq m}\}$ where the sequence $(\tau_i)_{1 \leq i \leq m}$ is nondecreasing, and ℓ_i denotes label of the option which is exercised at τ_i. In this section, we allow agents to exercise options which are out-of-the-money (i.e., pay $K^{(i)} > X_{\tau_i}$ to receive cash value X_{τ_i}). However as we soon argue, such a strategy is clearly suboptimal. Furthermore, we allow agents to follow a strategy which involves never exercising an option, in which case it is discarded unexercised. This explains why we may have $m < n$. We say that such an exercise strategy is feasible if it respects the vesting and maturity requirements, so that if $\ell_i = l$ then $t \vee V^{(l)} \leq \tau_i \leq T^{(l)}$.

Now we discuss the optimality, or otherwise, of various exercise strategies. The first result is completely natural. The second depends on Hypothesis 3.5 (prefer more to less and cash now rather than later), and on an ordering property for the characteristics of the options.

PROPOSITION 6.1. *Suppose Hypothesis 3.1 holds. Any exercise strategy which involves paying $K^{(i)}$ to receive X_{τ_i} when $K^{(i)} > X_{\tau_i}$ is suboptimal. Any exercise strategy which leaves unexercised options which expire in the money is suboptimal.*

Consider a portfolio of options with characteristics $(K^{(i)}, V^{(i)}, T^{(i)})_{1 \leq i \leq n}$. We say that the strike, vesting date, and maturity are comonotonic if there is a relabeling of the options, represented by a permutation σ of the labels, such that

$$K^{(\sigma(i))} \leq K^{(\sigma(j))}; \qquad T^{(\sigma(i))} \leq T^{(\sigma(j))}; \qquad V^{(\sigma(i))} \leq V^{(\sigma(j))}.$$

THEOREM 6.2. *Suppose Hypothesis 3.5 holds. Suppose that the characteristics of the options are such that the strike, vesting date, and maturity are comonotonic. Then, for any exercise strategy, there is a modified exercise strategy in which the options are exercised in order of increasing strike (and then also maturity and vesting date) for which the value of the option portfolio is at least as large as the value under the original exercise strategy.*

COROLLARY 6.3. *Suppose Hypothesis 3.5 holds, and that the characteristics of the options are such that the strike, vesting date, and maturity are comonotonic. Then, in searching for optimal strategies, it is sufficient to look in the class of strategies for which no option is exercised unless it is in the money, every option is exercised if it has reached maturity (and is discarded unless it is in the money), and options are exercised in order of increasing strike.*

The proof of the theorem shows that any exercise strategy may be improved upon if on the same set of exercise dates, the options are instead exercised in order of increasing strike/maturity/vesting date. Such an ordered strategy generates at least as much wealth as the original strategy—we show this by showing the cumulative amount spent on strikes at each exercise date is never more (and could be less) for the ordered strategy. The key point is to show that for the ordered strategy, it is always feasible to exercise the relevant option—that is, that it has vested and has not expired. This is where the comonotonicity is important.

Proof of Theorem 6.2. Without loss of generality we may assume that the options are labelled such that σ is the identity permutation and then for $i < j$,

$$K^{(i)} \leq K^{(j)} \qquad T^{(i)} \leq T^{(j)} \qquad V^{(i)} \leq V^{(j)}.$$

Fix an element of the sample space ω. Consider a first (male) agent, and suppose he follows the strategy $\mathcal{S}^M = \{(\tau_j^M, \ell_j^M)_{1\leq j \leq m}\}$, with resultant cashflow

$$I_s^{(t),M} \equiv I_s^{(t),M}\left(\left\{\left(\tau_j^M, \ell_j^M\right)\right\}\right) = \sum_{j:\tau_j^M \leq s} \left(X_{\tau_j^M} - K^{(\ell_j^M)}\right).$$

Now consider a second agent (who we take to be female). Suppose this agent exercises options on the same dates as the male agent. In each case she exercises the option with lowest label which has vested, but not yet expired. (We argue below that this set is nonempty, so that there is an option she can exercise.) Write her strategy as $\mathcal{S}^F = \{(\tau_i^F, \ell_i^F)_{1\leq i \leq m}\}$; then $\tau_i^F = \tau_i^M$, though the labels (ℓ_i^M, ℓ_i^F) may be different. Her cashflow is

$$I_s^{(t),F} \equiv I_s^{(t),F}\left(\left\{\left(\tau_i^F, \ell_i^F\right)\right\}\right) = \sum_{i:\tau_i^F \leq s} (X_{\tau_i^F} - K^{(\ell_i^F)}).$$

Suppose that we can show that for each j,

$$\sum_{i\le j} K^{(\ell_i^M)} \ge \sum_{i\le j} K^{(\ell_i^F)}. \tag{6.1}$$

Then if $j = \sup\{k : \tau_k^M \le s\}$,

$$\begin{aligned} I_s^{(t),M} &= \sum_{i\le j}\left(X_{\tau_i^M} - K^{(\ell_i^M)}\right) \\ &\le \sum_{i\le j} X_{\tau_i^F} - \sum_{i\le j} K^{(\ell_i^F)} = I_s^{(t),F}. \end{aligned}$$

Now suppose that the two agents have identical preferences, and give identical valuations to cashflows, or rather suppose that we are considering the choices of a single agent between cashflows. Then, given the cashflows are ordered, under Hypothesis 3.5 we have $\mathcal{V}(I^{(t),M}) \le \mathcal{V}(I^{(t),F})$, and that the agent does no worse, and may do better by exercising the options in order of increasing strike. It remains to prove (6.1).

Given an m-tuple $\gamma = (\gamma_1, \dots, \gamma_m)$ of distinct labels ($\gamma_i \in \{1, \dots, n\}$) we can define the ordered m-tuple $\tilde{\gamma} = (\gamma_{(1)}, \dots, \gamma_{(m)})$ where

$$\gamma_{(1)} = \min\{\gamma_1, \dots, \gamma_m\} \qquad \gamma_{(k)} = \min\{k \in \{\gamma_1, \dots, \gamma_m\} \setminus \{\gamma_{(1)}, \dots, \gamma_{(k-1)}\}\}$$

Now, given two m-tuples γ, δ we can define a partial order via $\gamma \preceq_m \delta$ if $\gamma_{(i)} \le \delta_{(i)}$ for each $i \le m$. We want to show that for each j, $\ell^F := (\ell_i^F)_{1\le i\le j} \preceq_j (\ell_i^M)_{1\le i\le j} =: \ell^M$, then (6.1) follows easily from the monotonicity of the sequence $K^{(i)}$.

Note that, by construction, the elements ℓ_i^F are increasing, so that $\tilde{\ell}^F = \ell^F$. (If that were not the case, then we would have $\ell_k^F < \ell_i^F$ for some $k > i$. Then the option with label ℓ_k^F would have vested by the date τ_i^F, since the the option with label ℓ_i^F has vested; moreover this option cannot have matured, since it has still not matured by date τ_k^F. Hence this option was available to be chosen by the female agent on date τ_i^F, contradicting the assumption that the female agent exercises the option with smallest label.)

Fix $j \le m$, and recall $\tilde{\ell}^M$ is the ordered set of labels corresponding to $(\ell_i^M)_{1\le i\le j}$. For each $k \le j$ we want to show that $\tilde{\ell}_k^M \equiv \ell_{(k)}^M \ge \ell_k^F$. Suppose this is true for $k \le r-1$. If we can show that $\tilde{\ell}_r^M > \ell_{r-1}^F$, and that $V^{(\tilde{\ell}_r^M)} \le \tau_r^F \le T^{(\tilde{\ell}_r^M)}$, then the option with label $\tilde{\ell}_r^M$ has vested, but not yet matured on date τ_r^F, and has not yet been exercised by the female agent. Hence, this option is available to be exercised by the female agent, and since she exercises the option with smallest label, $\ell_r^F \le \tilde{\ell}_r^M$ as required. But $\tilde{\ell}_r^M > \tilde{\ell}_{r-1}^M \ge \ell_{r-1}^F$ by the inductive hypothesis. Moreover, note that for any subset $\mathcal{L} \subseteq \{\ell_i^M : 1 \le i \le j\}$ of size r, $\max\{l \in \mathcal{L}\} \ge \max\{\tilde{\ell}_i^M : 1 \le i \le r\} = \tilde{\ell}_r^M$. Then, since the male agent has exercised all the options with labels $\ell_i^M : 1 \le i \le r$ by date $\tau_r^M = \tau_r^F$,

$$\tau_r^F \ge \max\left\{V^{(l)}; l \in \left\{\ell_i^M : 1 \le i \le r\right\}\right\} \ge \max\left\{V^{(l)}; l \in \left\{\tilde{\ell}_i^M : 1 \le i \le r\right\}\right\} = V^{(\tilde{\ell}_r^M)}$$

and the option with label $\tilde{\ell}_r^M$ has vested. Similarly,

$$\tau_r^F \le \min\left\{T^{(l)}; l \in \left\{\ell_i^M : r \le i \le j\right\}\right\} \le \min\left\{T^{(l)}; l \in \left\{\tilde{\ell}_i^M : r \le i \le j\right\}\right\} = T^{(\tilde{\ell}_r^M)}$$

and the option with label $\tilde{\ell}_r^M$ has not yet matured. □

Although Theorem 6.2 is stated for fixed maturities, since the proof fixes ω, it also holds for ordered but random maturities $\mathcal{T}^{(1)}(\omega) \le \dots \mathcal{T}^{i-1}(\omega) \le \mathcal{T}^{i}(\omega) \le \cdots \le \mathcal{T}^{(n)}(\omega)$.

This is useful in the context of employment termination. In this case we could set the maturities to be $\mathcal{T}^{(i)} = T^{(i)} \wedge \tau^{\gamma}$, where τ^{γ} is an independent random time (e.g., exponentially distributed with intensity γ) representing the termination time. Employees with stock options may leave their employment either voluntarily or nonvoluntarily, retire from their position, or die. In each of these cases, although the legal terms may differ, typically, nonvested options are cancelled, vested out-of-the-money options are cancelled, and vested in-the-money options must be exercised, perhaps within a short window of time. A second interpretation of τ^{γ} could be the time of an exogenous income shock which forces the agent to exercise her portfolio for liquidity reasons. Again, the conclusion is that optimal strategies are associated with exercising in label order of increasing characteristics.

6.1. Utility Indifference Pricing

We return to the model given in Section 5, here the employee's goal is to maximize expected utility of terminal wealth at some future date T ($T \geq \max(T^{(i)})$). Given Proposition 5.1, it is immediate that the conclusions of Theorem 6.2 hold. Thus, in the utility indifference model, under comonotonicity of strikes and maturities/vesting dates, we know it is optimal to exercise in order of increasing strike.

Now assume that the cashflow $I^{(t)}$ is the income from exercise of options in the portfolio. For an exercise strategy $\mathcal{S} = (\tau_i, l_i)_{1\leq i}$ we write $W = W^{\theta,\mathcal{S}} = W^{\theta, I^{(t)}(\mathcal{S})}$ where

$$I_u^{(t)}(\mathcal{S}) = \sum_{\tau_i \leq u} (X_{\tau_i} - K^{(i)})$$

provided $V^{(l_i)} \leq \tau_i \leq T^{(l_i)}$ and $\mathcal{S}$ is feasible.

The utility value of the stream of income $I^{(t)}$ is given by

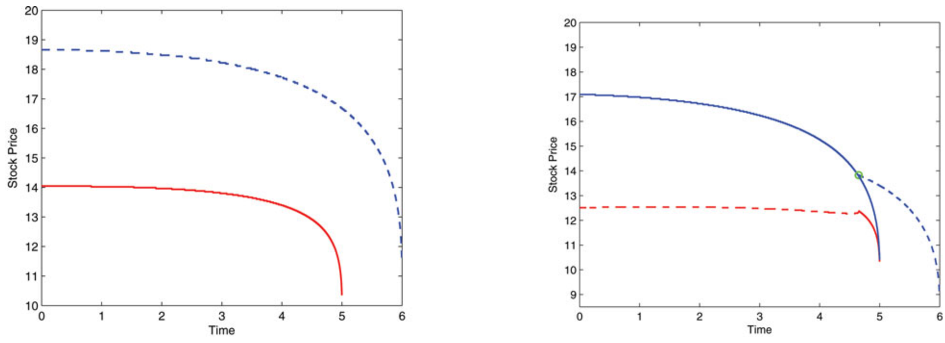
$$\mathcal{V}(\mathcal{S}) = \mathcal{V}(w, \mathcal{S}) = \mathcal{V}\left(w, I_{\mathcal{S}}^{(t)}\right) = \bar{U}^{-1}\Big(\sup_{\theta} \mathbb{E}\big[U\big(W_T^{\theta,\mathcal{S}}\big) | W_t = w,\, X_t = x\big]\Big) - w.$$

Optimizing over exercise strategies, the utility value of the portfolio of American options is

$$\mathcal{A} = \mathcal{A}(w, x, t) = \bar{U}^{-1}\Big(\sup_{\theta} \sup_{\mathcal{S}} \mathbb{E}\big[U\big(W_T^{\theta,\mathcal{S}}\big) | W_t = w,\, X_t = x\big]\Big) - w. \tag{6.2}$$

These definitions are analogous to those given earlier for the single option and this portfolio value can be calculated by dynamic programming.

The results of Theorem 6.2 greatly simplify the analysis of this problem. In general, if there is no comonotonicity, we need to consider each label order $(l_1, l_2, \ldots)$. Once we have solved for the optimal hedge and the exercise times $\tau_1, \tau_2, \ldots$ we can do a final optimization over the labels (exercise order). This involves solving up to $n!$ separate optimization problems (each of which itself involves optimizing over hedging strategies and exercise times, conditional on exercising in a particular order) and then finally optimizing over label order. However, in contrast, if K, V, T are comonotonic, then we know it is optimal to exercise in label order, reducing the problem to a single dynamic programming exercise. Firms often grant employee stock options which are at-the-money and with a fixed vesting date of say 3 years. If such options are granted in a bull market, then their characteristics will indeed be comonotonic.



(a) Co-monotonic: $(K^{(2)} = 11, T^{(2)} = 6, V^{(2)} = 0)$

(b) $(K^{(2)} = 8.5, T^{(2)} = 6, V^{(2)} = 0)$

FIGURE 6.1. Exercise thresholds for a portfolio of two options calculated from (6.2) with $U(x) = -e^{-\gamma x}$ via dynamic programming. The solid line gives the threshold for call with $(K^{(1)} = 10,\ T^{(1)} = 5,\ V^{(1)} = 0)$. The dashed line in each panel is the threshold for the call with $(T^{(2)} = 6,\ V^{(2)} = 0)$ and different strike in each panel as indicated. Other parameter values are: $\gamma = 0.1$, $\rho = 0.5$, $r = 0.05$, $q = 0.02$, $\sigma = 0.4$, $T = 6 = \max(T^{(1)}, T^{(2)})$.

6.2. Example—CARA Utility

We assume utility is CARA, $U(x) = -e^{-\gamma x}; \gamma > 0$. We will solve the dynamic programming problem resulting from the valuation in (6.2). The resulting free boundary problem is described in the Appendix (Section (ii)) together with a brief description of the numerical scheme. We illustrate the optimal exercise boundaries for a portfolio of two options in Figure 6.1.

In each panel in the figure, the solid line is the exercise threshold for the call with characteristics $(K^{(1)} = 10,\ T^{(1)} = 5,\ V^{(1)} = 0)$. In panel (a), the dashed line is the exercise threshold for call with $(K^{(2)} = 11,\ T^{(2)} = 6,\ V^{(2)} = 0)$, so the characteristics are comonotonic. We see that as Theorem 6.2 predicts, the low strike/maturity option is exercised first, as it has the lower threshold. Comparing to the stand-alone thresholds in panel (a) of Figure 5.1, we see that when treated as a portfolio, the threshold for the shorter maturity option has moved downwards, shrinking the continuation region for the shorter maturity option. This is because the presence of the longer maturity option in the agent's portfolio causes her to exercise the shorter maturity option at lower price levels, in order to unwind some risk. The longer maturity threshold remains the same as the stand-alone threshold in Figure 5.1—because once the shorter maturity option is exercised the agent only has the longer maturity option in her portfolio.

In panel (b), the strikes and maturities are not comonotonic. The dashed line is the threshold for the call with $(K^{(2)} = 8.5,\ T^{(2)} = 6,\ V^{(2)} = 0)$. The circle (at time 4.65 years) indicates the time at which the first option to be exercised switches from the longer maturity to the shorter maturity option. Starting below the lower boundary, one possibility is that we hit the dashed threshold (before 4.65 years) and exercise the $T = 6$ option first. If so, the relevant boundary for the $T = 5$ year option is the solid line (which continues to 5 years and is exactly the stand-alone boundary for the $T = 5$ option given in Figure 5.1). If, instead, the price hits the solid piece of the lower boundary (which only exists between 4.65 and 5 years), then the $T = 5$ option is exercised first and the higher solid line over the period (4.65,5) is redundant. Then the relevant boundary for the

$T = 6$ option is the dashed line between 4.65 and 6 years. Again, this is the stand-alone boundary for the $T = 6$ option, since once the shorter maturity option is exercised, only one option remains.

This example extends the literature on utility-based employee stock option models, which previously studied such problems for single (or identical) options. As we see in panel (b), if the strikes and maturities are not comonotonic, either option could be exercised first, and there are many factors which will influence the precise positioning of the boundaries, and in particular, where they intersect.

Even more complex is the situation with CRRA utility where the computations involve an additional wealth dimension. Carpenter et al. (2010) studied American call exercise in the CRRA utility indifference model for a single option. Given the additional wealth dimension, their problem was already numerically challenging with three state variables (time, stock price, and wealth). Again, of course, our exercise ordering result holds and will reduce the computational burden for portfolios.

7. CONCLUSIONS

The paper aims to study the optimal exercise of American options in a setting with minimal assumptions on the agent's preferences, valuation methodology, and prices. Our main result tells us that a portfolio of American calls with comonotonic strikes, maturities, and vesting dates should be exercised in order of increasing strike. Since employees often receive regular grants of American call options, they should exercise the lower strike options with the least time-to-go before the higher strike options with more time-to-go. Although we concentrate on the American call, similar results will hold for the put. A version of Theorem 6.2 will hold with the proviso that strikes and maturities/vesting dates are counter comonotonic, and puts will be exercised in increasing order of maturity but decreasing strike.

We illustrate the portfolio exercise result in a standard CARA utility indifference model and thus give the first treatment of portfolios of options with different characteristics in this setting. Interestingly, our portfolio exercise ordering result can be thought of as a generalization of the single option result of Theorem 3.7. However, the portfolio result does not require that the valuation methodology has the monetary property, and thus holds more widely.

The strength of our results is that we require very few assumptions on preferences and prices. Rather than attempting to solve numerically for thresholds (which depend on preferences and the model for the asset price), we instead ask what can be shown concerning exercise ordering in the absence of such assumptions? This is advantageous for several reasons. It is empirically difficult to ascertain what the true underlying process is, and hence it is useful to develop results which are robust to different price specifications. Although some of our results rely on a diffusion assumption, Theorem 6.2 does not, and thus holds, for example, for models with jumps in prices such as Lévy processes. Second, as Section 4 shows, even under the assumption of risk neutral pricing, exercise and continuation regions can be complex. In incomplete markets (for example, under utility indifference) the regions can be even more complicated, and difficult to characterize, hence the more we can say about how regions "nest" or equivalently, the exercise ordering of a set of options—without having to construct explicitly the thresholds—the better. Finally, there is no consensus on the most appropriate preferences to model agent's decisions concerning risky outcomes, and our results stand under most popular choices. Although our examples focus on utility indifference pricing, our requirements on preferences are much weaker, and are valid for the S shaped function of prospect theory (Kahneman and Tversky 1979), or valuation using hyperbolic discounting.

APPENDIX: SOLVING THE FREE BOUNDARY PROBLEMS

A.1. (i) Black–Scholes Model

We briefly outline our solution approach for the Black–Scholes model used to produce the exercise boundaries in Figures 3.1, 4.1, and 4.2. Consider a single option with strike K, maturity T, and vesting date V, and assume we price under the Black–Scholes model. The value $V^{BS}(u, X_u)$ of holding the American option solves the well-known linear complementarity problem:

$$V^{BS}(u, X_u) \geq (X_u - K)^+$$

$$\dot{V}^{BS} + \mathcal{L}_{BS} V^{BS} - r V^{BS} \leq 0$$

and

$$\left(\dot{V}^{BS} + \mathcal{L}_{BS} V^{BS} - r V^{BS}\right)\left(V^{BS}(u, X_u) - (X_u - K)^+\right) = 0,$$

where the differential operator $\mathcal{L}_{BS}$ is defined by

$$\mathcal{L}_{BS} = \frac{\sigma^2 x^2}{2} \frac{\partial^2}{\partial x^2} + (r - q)x \frac{\partial}{\partial x}.$$

Boundary conditions are given by $V^{BS}(u, 0) = 0$ and $V^{BS}(T, X_T) = (X_T - K)^+$. The optimal exercise time τ_{BS} is defined by

$$\tau_{BS} = \inf\left\{t \vee V \leq u \leq T : V^{BS}(u, X_u) = (X_u - K)^+\right\}$$

which defines an exercise boundary

$$X^*_{BS}(u) = \inf\left\{x \geq 0 : V^{BS}(u, x) = (x - K)^+; u \in [t \vee V, T]\right\}$$

and thus

$$\tau_{BS} = \inf\{t \vee V \leq u \leq T : X_u = X^*_{BS}(u)\}.$$

We use a Crank–Nicolson finite difference method to solve the PDE, with a projected successive over relaxation algorithm (PSOR) to enforce the free boundary constraint. For similar schemes, see Wilmott, Howison, and Dewynne (2005).

A.2. (ii) Portfolio of Options

We describe our dynamic programming approach for the utility indifference pricing of the portfolio of American options in (6.2). This approach is used to produce the exercise boundaries in Figures 5.1, 5.2, and 6.1.

The portfolio of n options have ordered maturities $T^{(i)} < T^{(j)}$; $1 \leq i < j \leq n$ and strikes $K^{(i)}$; $i = 1, \ldots, n$. We take $V^{(i)} = 0$; $i = 1, \ldots, n$ for simplicity here. The goal is to maximize expected utility of terminal wealth at a future date $T \geq T^{(n)}$.

Recall from (5.3) that the employee's wealth $W = W^{\theta, I^{(i)}}$ consists of two components—her trading wealth from investment in the market asset and the cashflow $I^{(i)}$. Denote her trading wealth by Y, defined by $Y = W^{\theta,0}$. The dynamics of Y are

$$dY_u = \big(rY_u + \theta_u(\mu^M - r)\big)\,du + \theta_u\sigma^M dZ_u; \quad Y_t = y = w.$$

We also recall the concept of the indirect utility for terminal wealth obtained without any income, $\bar{U}(w, t)$, which was defined in Section 5 and satisfies

$$\bar{U}(w, t) = \bar{U}(y, t) = \bar{U}(y) = \sup_{\{\theta_s\}_{t\le s\le T}} \mathbb{E}\,U(Y_T | Y_t = y).$$

Denote by Π the portfolio of unexercised options and by $|\Pi|$ the number of remaining options. For example, if options with labels j, k remain unexercised then $\Pi = \{j, k\}$ and $|\Pi| = 2$. Define also the shortest maturity left in the portfolio by $T_\Pi = min\{T^{(i)}: i \in \Pi\}$.

The value to the executive $V^\Pi(u, Y_u, X_u)$ with remaining options Π, current trading wealth Y_u, and current stock price X_u solves the following linear complementarity problem:

$$V^\Pi(u, Y_u, X_u) \ge \max_{i\in\Pi}\{V^{\Pi\setminus\{i\}}(u, Y_u + (X_u - K^{(i)})^+, X_u)\} \tag{A.1}$$

$$\dot{V}^\Pi + \sup_\theta\{\mathcal{L}V^\Pi\} \le 0 \tag{A.2}$$

and

$$\left(\dot{V}^\Pi + \sup_\theta\{\mathcal{L}V^\Pi\}\right)\left(\max_{i\in\Pi}\{V^{\Pi\setminus\{i\}}(u, Y_u + (X_u - K^{(i)})^+, X_u) - V^\Pi(u, Y_u, X_u)\right) = 0,$$

where the differential operator $\mathcal{L}$ is defined by

$$\mathcal{L} = \frac{\sigma^2x^2}{2}\frac{\partial^2}{\partial x^2} + (\nu - q)x\frac{\partial}{\partial x} + \rho\theta\sigma^M\sigma x\frac{\partial^2}{\partial y\partial x} + \frac{\theta^2(\sigma^M)^2}{2}\frac{\partial^2}{\partial y^2} + [\theta(\mu^M - r) + ry]\frac{\partial}{\partial y}.$$

Boundary conditions are given by $V^\Pi(u, Y_u, 0) = \bar{U}(Y_u; u)$ and

$$\forall\{i\} \in \Pi; \quad V^\Pi(T^{(i)}, Y_{T^{(i)}}, X_{T^{(i)}}) = V^{\Pi\setminus\{i\}}(T^{(i)}, Y_{T^{(i)}} + (X_{T^{(i)}} - K^{(i)})^+, X_{T^{(i)}}).$$

The optimal exercise times $\tau_{[n]} \le \ldots \le \tau_{[1]}$, where $\tau_{[i]}$ is the exercise time at which there are i options *remaining* in the portfolio (contrast to the increasing set of τ_i in Section 6), are defined by

$$\tau_{[|\Pi|]} = \inf\{t \le u \le T_\Pi : V^\Pi(u, Y_u, X_u) = \max_{i\in\Pi}\{V^{\Pi\setminus\{i\}}(u, Y_u + (X_u - K^{(i)})^+, X_u)\}\}.$$

When only one option remains, $|\Pi| = 1$ and the optimal exercise time is given by

$$\tau_{[1]} = \inf\{t \le u \le T_\pi : V^\Pi(u, Y_u, X_u) = V^\emptyset(u, Y_u + (X_u - K^{(i)})^+, X_u)\},$$

where $V^\emptyset(u, Y_u, X_u) = \bar{U}(Y_u; u)$. Optimization gives the hedging strategy

$$\theta_u = (-(\mu^M - r)V^\Pi_y - xV^\Pi_{yx}\rho\sigma^M\sigma)/((\sigma^M)^2V^\Pi_{yy}); \quad u \le \tau_{[|\Pi|]}$$

which is substituted into the PDE.

A.2.1. *(a) CARA utility.* For $U(x) = -e^{-\gamma x}$ we can compute the indirect utility for terminal wealth as

$$\bar{U}(y;t) = -e^{-\gamma y e^{r(T-t)}} e^{-\frac{(\mu^M - r)^2}{(\sigma^M)^2}(T-t)}.$$

Under CARA utility, we can use separation of variables and a power transformation

$$V^\Pi(u, y, x) = \bar{U}(y;u) H^\Pi(u, x)^{1/(1-\rho^2)}$$

to remove the wealth state variable. The complementarity problem and boundary conditions can be restated in terms of the new variable $H^\Pi(u, x)$. Associated with each $H^\Pi(u, X_u)$ there is a free boundary

$$X^*_{[|\Pi|]}(u) = \inf\{x \geq 0 : H^\Pi(u, x) = \min_{i \in \Pi}\{e^{-\gamma(1-\rho^2)(x-K^{(i)})^+ e^{r(T-u)}} H^{\Pi \setminus \{i\}}(u, x)\}; u \in [t, T_\Pi]\}$$

which represents the exercise boundary for the next option when the options Π remain, and the optimal exercise times can be represented as

$$\tau_{[|\Pi|]} = \inf\{t \leq u \leq T_\Pi : X_u = X^*_{[|\Pi|]}(u)\}.$$

We again use a Crank Nicolson finite difference method to solve the PDE, with a projected successive over relaxation algorithm to to enforce the free boundary constraint. Figure 6.1 depicts the resulting exercise boundaries for a portfolio of two options. Figure 5.1 represents the exercise boundaries for a single option, $|\Pi| = 1$.

Taking $\gamma \to 0$ or $|\rho| \to 1$ in the above utility indifference pricing algorithm will recover the Black–Scholes values described in (i).

A.2.2. *(b) Disconnected exercise regions in Figure 5.2.* Take $|\Pi| = 1$ option and take zero investment in the market asset $\theta = 0$. Then $dY_u = rY_u du$; $Y_t = y$ and for $U(x) = \frac{x^{1-A}}{1-A} + cx$,

$$\bar{U}(y) = \bar{U}(y;t) = U(Y_T | Y_t = y).$$

The value to the employee of holding the American option with strike K and maturity T is given by

$$V(u, X_u) = \sup_{t \leq \tau \leq T} \mathbb{E}\, U((X_\tau - K)^+ e^{r(T-\tau)} + Y_T)$$

and solves

$$V(u, X_u) \geq \bar{U}((X_u - K)^+ + y, u)$$

$$\dot{V} + \mathcal{L}V \leq 0,$$

where

$$\mathcal{L} = \frac{\sigma^2 x^2}{2} \frac{\partial^2}{\partial x^2} + (r - q)x \frac{\partial}{\partial x}.$$

Boundary conditions are given by $V(u,0) = \bar{U}(Y_u)$ and $V(T, X_T) = \bar{U}(Y_T + (X_T - K)^+, T)$. The optimal exercise time τ is defined by

$$\tau = \inf\{t \le u \le T : V(u, X_u) = \bar{U}((X_u - K)^+ + y, u)\}.$$

Again, we solve the free boundary problem using a similar numerical scheme as described earlier. Figure 5.2 depicts the resulting exercise boundaries. Carpenter et al. (2010) describe a scheme to solve a similar example (without dividends) in their section 4.

REFERENCES

BABILUA, P., I. BOKUCHAVA, B. DOCHVIRI, and M. SHASHIASHVILI (2007): The American Put Option in a One-Dimensional Diffusion Model with Level-Dependent Volatility, *Stochastics* 79(1–2), 5–25.

BAYRAKTAR, E., I. KARATZAS, and S. YAO (2010): Optimal Stopping for Dynamic Convex Risk Measures, *Ill. J. Math.* 54(3), 1025–1067.

BAYRAKTAR, E., and S. YAO (2011): Optimal Stopping for Non-linear Expectations—Parts I and II, *Stoch. Process. Their Applicat.* 121, 185–264.

BENDER, C. (2011): Dual Pricing of Multi-Exercise Options under Volume Constraints, *Finance Stoch.* 15(1), 1–26.

BENSOUSSAN, A. (1984): On the Theory of Option Pricing, *Acta Appl. Math.* 2(2), 139–158.

BERGMAN, Y. Z., B. D. GRUNDY, and Z. WIENER (1996) General Properties of Option Prices, *J. Finance* 51(5), 1573–1610.

CARMONA, R., ed. (2008): *Indifference Pricing: Theory and Applications*, Princeton, NJ: Princeton University Press.

CARMONA, R. and TOUZI, N. (2008): Optimal Multiple Stopping and Valuation of Swing Options, *Math. Finance* 18(2), 239–268.

CARPENTER, J. N., R. STANTON, and N. WALLACE (2010): Optimal Exercise of Executive Stock Options and Implications for Firm Cost, *J. Finan. Econ.* 98, 315–337.

COX, J. C., and M. RUBINSTEIN (1985): *Options Markets*, Englewood Cliffs, NJ: Prentice Hall Inc.

CVITANIC, J., Z. WIENER, and F. ZAPATERO (2008): Analytic Pricing of Employee Stock Options, *Rev. Finan. Stud.* 21, 683–724.

DAVIS, M. H. A. (1999): Option Valuation and Hedging with Basis Risk, in *System Theory: Modeling, Analysis and Control*, T.E. Djaferis and I.C. Schick, eds., Amsterdam: Kluwer, pp. 245–254.

EKSTRÖM, E. (2004): Properties of American Option Prices, *Stochastic Processes and Their Applications* 114, 265–278.

EL KAROUI, N., M. JEANBLANC-PICQUE, and S. E. SHREVE (1998): Robustness of the Black–Scholes Formula, *Math. Finance* 8(2), 93–126.

FÖLLMER, H., and A. SCHIED (2004): *Stochastic Finance: An Introduction in Discrete Time*, 2nd revised and extended edition, de Gruyter Studies in Mathematics 27, Berlin: Walter de Gruyter and Co.

FREI, C., and M. SCHWEIZER (2008): Exponential Utility Indifference Valuation in Two Brownian Settings with Stochastic Correlation, *Adv. Appl. Probab.* 40, 401–423.

GÖTTSCHE, O. E., and M. H. VELLEKOOP (2011): The Early Exercise Premium for the American Put under Discrete Dividends, *Math. Finance* 21(2), 335–354.

GRASSELLI, M., and V. HENDERSON (2009): Risk Aversion and Block Exercise of Executive Stock Options, *J. Econ. Dyn. Contl.* 33, 109–127.

HENDERSON, V. (2005): The Impact of the Market Portfolio on the Valuation, Incentives and Optimality of Executive Stock Options, *Quant. Finance* 5(1), 1–13.

HENDERSON, V. (2007): Valuing the Option to Invest in an Incomplete Market, *Math. Finan. Econ.* 1(2), 103–128.

HENDERSON, V., and D. HOBSON (2008): Utility Indifference Pricing: An Oveview, in *Indifference Pricing: Theory and Applications*, R. Carmona, ed., Princeton NJ: Princeton University Press, pp. 44–72.

HENDERSON, V., and D. HOBSON (2011): Optimal Liquidation of Derivative Portfolios, *Math. Finance* 21(3), 365–382.

HENDERSON, V., and J. SUN (2010): Employee Stock Options, *Encyclopaedia of Quantitative Finance*, R. Cont, ed., Chichester, UK: John Wiley and Sons Ltd., pp. 561–566.

HOBSON, D. (1998): Volatility Misspecification, Option Pricing and Superreplication via Coupling, *Ann. Appl. Probab.* 8(1), 193–205.

IBÁÑEZ, A. (2004): Valuation by Simulation of Contingent Claims with Multiple Early Exercise Opportunities, *Math. Finance* 14, 223–248.

JACKA, S. D. (1991): Optimal Stopping and the American Put, *Math. Finance* 1, 1–14.

KAHNEMAN, D., and A. TVERSKY (1979): Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 46, 171–185.

KARATZAS, I. (1988): On the Pricing of American Options, *Appl. Math. Optim.* 17, 37–60.

KARATZAS, I., and S. E. SHREVE (1994): *Brownian Motion and Stochastic Calculus*, New York: Springer Verlag.

KIM, I. J. (1990): The Analytic Valuation of American Options, *Rev. Finan. Stud.* 3, 542–572.

KRÄTSCHMER, V., and J. SCHOENMAKERS (2010): Representations for Optimal Stopping under Dynamic Monetary Utility Functionals, *SIAM J. Finan. Math.* 1, 811–832.

LEUNG, T., and R. SIRCAR (2009): Accounting for Risk Aversion, Vesting, Job Termination Risk and Multiple Exercises in Valuation of Employee Stock Options, *Math. Finance*, 19, 99–128.

LINDVALL, T. (1992): *Lectures on the Coupling Method*, New York: Wiley.

MCKEAN, H. P. (1965): A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics, *Ind. Manage. Rev.* 6, 32–39.

MEINSHAUSEN, N., and B. M. HAMBLY (2004): Monte Carlo Methods for the Valuation of Multiple Exercise Options, *Math. Finance* 14, 557–583.

MERTON, R. C. (1973): Theory of Rational Option Pricing, *Bell J. Econ. Manage. Sci.* 4, 141–183.

MYNENI, R. (1992): The Pricing of the American Option, *Ann. Appl. Probab.* 2, 1–23.

RIEDEL, F. (2009): Optimal Stopping with Multiple Priors, *Econometrica* 77(3), 857–908.

ROGERS, L. C. G., and J. SCHEINKMAN (2007): Optimal Exercise of Executive Stock Options, *Finance Stoch.* 11, 357–372.

ROGERS, L. C. G., and D. WILLIAMS (2000): *Diffusions, Markov Processes and Martingales: Vol. 2, Itô Calculus*, 2nd ed., Cambridge, UK: Cambridge University Press.

VAN MOERBEKE, P. (1976): On Optimal Stopping and Free Boundary Problems, *Arch. Rational Mech. Anal.* 60, 101–148.

VELLEKOOP, M. H., and J. W. NIEUWENHUIS (2009): An Integral Equation for American Put Options on Assets with General Dividend Processes, Preprint.

WILMOTT, P., S. HOWISON, and J. DEWYNNE (2005): *The Mathematics of Financial Derivatives*, Cambridge, UK: Cambridge University Press.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 567–597

LIMIT THEOREMS FOR PARTIAL HEDGING UNDER TRANSACTION COSTS

YAN DOLINSKY

Department of Mathematics, ETH Zurich

We study shortfall risk minimization for American options with path-dependent payoffs under proportional transaction costs in the Black–Scholes (BS) model. We show that for this case the shortfall risk is a limit of similar terms in an appropriate sequence of binomial models. We also prove that in the continuous time BS model, for a given initial capital, there exists a portfolio strategy which minimizes the shortfall risk. In the absence of transactions costs (complete markets) similar limit theorems were obtained by Dolinsky and Kifer for game options. In the presence of transaction costs the markets are no longer complete and additional machinery is required. Shortfall risk minimization for American options under transaction costs was not studied before.

KEY WORDS: American options, shortfall risk, transaction costs.

1. INTRODUCTION

This paper deals with shortfall risk minimization for American options under proportional transaction costs. It is well known that in a complete market an American contingent claim can be hedged perfectly with an initial capital, which is equal to the optimal stopping value of the discounted payoff under the unique martingale measure. In the presence of transaction costs, the market is no longer complete and the initial capital required for perfect hedging (superhedging price) of the options is often too high. Several authors (see, for example, Soner, Shreve, and Cvitanic 1995; Levental and Skorohod 1997; Cvitanic, Pham, and Touzi 1999) showed that the superhedging price of European call options (also of American call options) in the Black–Scholes (BS) model is equal to the price of the least expensive buy-and-hold (superhedging) strategy. In Jakubenas, Levental, and Ryznar (2003) these results were extended to path-dependent options. For example, it was demonstrated that for European and American options (in the BS model) with a Russian type of payoffs, the superhedging price is infinite, i.e., perfect hedging is not available. Therefore, with the presence of transaction costs, it is reasonable to assume that the seller's (investor's) initial capital is less than the superhedging price. In this case, the seller is ready to accept the risk that his portfolio value at an exercise time may be less than his obligation to pay and he will need additional funds to fulfill the contract. This leads to the natural question of minimization of risk for a given amount of initial capital. In order to make this question precise we need to define explicitly the risk measure.

We deal with a certain type of risk called the shortfall risk, which is defined for American options as the maximal expectation with respect to the buyer exercise times of the

I am grateful to the referees for many valuable suggestions, which lead to a substantial improvement of the paper.

Manuscript received May 2010; final revision received January 2012.

Address correspondence to Yan Dolinsky, Department of Mathematics, ETH, Zurich 8092, Switzerland; e-mail: yan.dolinsky@math.ethz.ch

DOI: 10.1111/j.1467-9965.2012.00525.x

discounted shortfall (see Mulinacci 2011). In the presence of transaction costs, the problem of shortfall risk minimization was studied only for European options (see Kamizono 2001, 2003; Guasoni 2002a,b; Trivellato 2009). The first two authors considered a general setup for which they proved that for a given initial capital there exists a portfolio strategy which minimizes the shortfall risk. In Trivellato (2009), shortfall risk minimization is studied for European options in a binomial model. It is shown that for a given initial capital, the shortfall risk and the corresponding optimal portfolio can be calculated by a dynamical programming algorithm.

In this paper we study shortfall risk minimization for cash-settled American options in the BS model. We consider path-dependent payoffs with some regularity conditions. We allow only self-financing portfolios, which satisfy the no-bankruptcy condition, i.e., portfolios with a nonnegative wealth process. This corresponds to the situation when the portfolio is handled without borrowing of capital. By using convexity of the shortfall risk measure, we show that for a given initial capital there exists a portfolio strategy which minimizes the risk. From a practical view point, existence results are not sufficient. An investor with a fixed initial capital wants to compute the minimal possible shortfall risk and to find explicitly a portfolio strategy that minimizes or "almost" minimizes the shortfall risk. For binomial models the above problems can be solved by a dynamical programming algorithm.

Our approach is to use an appropriate sequence of binomial models in order to approximate the shortfall risk and to construct "almost" optimal portfolios in the BS model. Our main results are the following. We show that under proportional transaction costs, the shortfall risk in the BS model is a limit of similar terms with the same proportional transaction costs in an appropriate sequence of binomial models. Furthermore, we use the optimal portfolios in the binomial models in order to construct "almost" optimal portfolios for the BS model. For the case where the payoff process is Lipschitz continuous we also provide error estimates for the above approximations.

Similar results were obtained in Dolinsky and Kifer (2008, 2011) for game options without the presence of transaction costs. The proof of the results there relied heavily on the completeness of the markets, which is no longer the case with the presence of transaction costs.

The main auxiliary result, which is crucial for proving the limit theorems in our setup is Lemma 4.2. This lemma provides a stability result for the shortfall risk as a function of the transaction costs parameters λ, μ. This result, together with the Skorohod embedding machinery, allows us to compare the shortfall risk in the BS model (under transaction costs) with the shortfall risks in the binomial models.

The paper is organized as follows. Main results of this paper are formulated in the next section. In Section 3 we analyze the binomial models and provide a dynamical programming algorithm for the shortfall risk and the corresponding optimal portfolios. In Section 4 we complete the proof of Theorems 2.2 and 2.3 (limit theorems), which are the main results of the paper. In Section 5 we prove Theorem 2.1, which provides an existence result for the optimal portfolio in the BS model. In Section 6 we show that in the BS model, when the transaction costs tend to 0, the corresponding shortfall risks converge to the shortfall risk in the complete BS market. Note that for the superhedging prices this is not true in general. For instance, the call option superhedging prices converge (as λ, $\mu \downarrow 0$) to the initial stock price, which is higher than the call option price in the complete BS market. The same occurs for American options with a Russian type of payoffs. In this case the limit of the superhedging prices (as λ, $\mu \downarrow 0$) is infinity. In Section 7 we provide a numerical study of the shortfall risk in the binomial models for call options.

For this case we study the behavior of the shortfall risk as a function of the initial capital. Furthermore, we analyze the dependence of these functions on the number of steps in the binomial models.

2. PRELIMINARIES AND MAIN RESULTS

Consider a complete probability space (Ω, P) together with a standard one-dimensional Brownian motion $\{W(t)\}_{t=0}^{\infty}$, and the filtration $\mathcal{F}_t = \sigma\{W(s) \mid s \le t\}$. We assume that the σ-algebras contain the null sets. A BS financial market consists of a savings account $B(t)$ with an interest rate r, assuming without loss of generality that $r = 0$, i.e.,

$$B(t) \equiv B_0 > 0, \tag{2.1}$$

and of a risky asset S, given by the equation

$$S(t) = S_0 \exp(\kappa W(t) + (\vartheta - \kappa^2/2)t), \quad S_0 > 0, \tag{2.2}$$

where $\kappa > 0$ is called volatility and $\vartheta \in \mathbb{R}$ is another constant denoting the drift. Denote by $\tilde{P}$ the unique martingale measure for the above model. Using standard arguments it follows that the restriction of the probability measure $\tilde{P}$ to the σ-algebra $\mathcal{F}_t$ satisfies

$$Z(t) := \frac{d\tilde{P}}{dP}\bigg| \mathcal{F}_t = \exp\left(-\frac{\vartheta}{\kappa} W(t) - \frac{1}{2}\left(\frac{\vartheta}{\kappa}\right)^2 t\right). \tag{2.3}$$

Let $T < \infty$ be the maturity date of our American option and let $\mathcal{T}_{[0,T]}$ be the set of all stopping times with respect to the filtration $\mathcal{F}$ which take values in $[0, T]$. Denote by $M[0, T]$ the space of all right continuous functions with left-hand limits (*càdlàg* functions). We consider the space $M[0, T]$ with the norm $||\upsilon|| = \sup_{0 \le t \le T} |\upsilon(t)|$. Let $\mathcal{C}(M)$ be the space of all continuous functions $F : [0, T] \times M[0, T] \to \mathbb{R}_+$ (with respect to the product topology) which satisfy the following conditions.

(i) There exists a constant $C > 0$ such that

$$\sup_{0 \le t \le T} F(t, \upsilon) \le C\Big(1 + \sup_{0 \le t \le T} |\upsilon(t)|\Big), \quad \forall \upsilon \in M[0, T]. \tag{2.4}$$

(ii) For all $t \in [0, T]$ and $\upsilon, \tilde{\upsilon} \in M[0, T]$, $F(t, \upsilon) = F(t, \tilde{\upsilon})$ if $\upsilon(s) = \tilde{\upsilon}(s)$ for all $s \le t$.

We define $\mathcal{C}_{Lip}(M) \subset \mathcal{C}(M)$ to be the set of all functions $F \in \mathcal{C}(M)$ for which there exists a constant L such that for any $t \ge s \ge 0$ and $\upsilon, \tilde{\upsilon} \in M[0, t]$,

$$\begin{gathered} |F(s, \upsilon) - F(s, \tilde{\upsilon})| \le L||\upsilon - \tilde{\upsilon}|| \quad \text{and} \\ |F(t, \upsilon) - F(s, \upsilon)| \le L\Big(|t - s|(1 + ||\upsilon||) + \sup_{u \in [s,t]} |\upsilon(t) - \upsilon(s)|\Big). \end{gathered} \tag{2.5}$$

Among examples of American options which fit into our setup are call or put options, Russian options, which are defined by

$$F(t, S) = \exp(-rt) \max\Big(M, \sup_{0 \le u \le t} \exp(ru) S(u)\Big),$$

and integral call or put options, which are defined by

$$F(t, S) = \exp(-rt)\left(\int_0^t \phi(\exp(ru)S(u))\,du - K\right)^+$$

or

$$F(t, S) = \exp(-rt)\left(K - \int_0^t \phi(\exp(ru)S(u))\,du\right)^+,$$

respectively. In the above expressions $r, M > 0$ are some positive constants and $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a Lipschitz continuous function. In all of the above cases $F \in \mathcal{C}_{Lip}(M)$. Since in our setup all terms are discounted, r can be interpreted as the real interest rate of the market.

Next, let $F \in \mathcal{C}(M)$ and consider a cash-settled American contingent claim with the payoff process given by

$$Y(t) = F(t, S), \quad 0 \le t \le T. \tag{2.6}$$

From the assumptions mentioned above it follows that $\{Y(t)\}_{t=0}^T$ is a continuous adapted stochastic process and $E\sup_{0\le t\le T} Y(t)$, $\tilde{E}\sup_{0\le t\le T} Y(t) < \infty$, where E and $\tilde{E}$ denote the expectations with respect to the probability measures P and $\tilde{P}$, respectively.

In our model, purchase and sale of the risky asset are subject to proportional transaction costs of rate λ and μ, respectively. We assume that $\lambda > 0$ and $0 < \mu < 1$ are constants. Thus, a trading strategy with a (finite) horizon T and an initial capital $x \ge 0$ is a pair $\pi = (x, \gamma)$, where $\gamma = \{\gamma(t)\}_{t=0}^T$ is an adapted process of bounded variation with left-continuous paths and $\gamma(0) = 0$. For any $t \in [0, T]$, $\gamma(t)$ is the number of stocks in the portfolio π at time t (before a transfer is made at this time). Set

$$\gamma^+(t) = \frac{\gamma(t) + \int_0^t |d\gamma(s)|}{2} \quad \text{and} \quad \gamma^-(t) = \frac{\int_0^t |d\gamma(s)| - \gamma(t)}{2}. \tag{2.7}$$

Clearly $\gamma(t) = \gamma^+(t) - \gamma^-(t)$ is a decomposition of γ into a positive variation γ^+ and a negative variation γ^-. The random variables $\gamma^+(t)$ and $\gamma^-(t)$, denote the cumulative number of stocks, purchased up to time t and sold up to time t (not including the transfers made at time t), respectively. The portfolio value (after liquidation) of a trading strategy π is given by

$$\begin{aligned} V^{\pi}_{\lambda,\mu}(t) = x &- (1+\lambda)\left(\int_0^t S(u)\,d\gamma^+(u) + \gamma(t)^- S(t)\right) \\ &+ (1-\mu)\left(\int_0^t S(u)\,d\gamma^-(u) + \gamma(t)^+ S(t)\right), \end{aligned} \tag{2.8}$$

where we denote $y^+ = \max(y, 0)$ and $y^- = \max(-y, 0)$. Observe that $V^{\pi}_{\lambda,\mu}(t)$ is the portfolio value at time t, before a transfer is made at this time. A self-financing strategy π is called *admissible* if the following no-bankruptcy condition holds

$$V^{\pi}_{\lambda,\mu}(t) \ge 0 \quad \forall t \in [0, T]. \tag{2.9}$$

The set of all *admissible* self-financing strategies with an initial capital x will be denoted by $\mathcal{A}(x, \lambda, \mu)$. For an *admissible* self-financing strategy π the shortfall risk is given by

$$R(\pi, \lambda, \mu) = \sup_{\tau \in \mathcal{T}_{[0,T]}} E\big[\big(Y(\tau) - V^{\pi}_{\lambda,\mu}(\tau)\big)^{+}\big], \tag{2.10}$$

which is the maximal possible expectation of the shortfall measured in cash. The shortfall risk for an initial capital x is given by

$$R(x, \lambda, \mu) = \inf_{\pi \in \mathcal{A}(x,\lambda,\mu)} \sup_{\tau \in \mathcal{T}_{[0,T]}} E\big[\big(Y(\tau) - V^{\pi}_{\lambda,\mu}(\tau)\big)^{+}\big], \tag{2.11}$$

i.e., $R(x, \lambda, \mu)$ is the infimum of all shortfall risks that can be achieved with initial capital x. A portfolio strategy $\pi \in \mathcal{A}(x, \lambda, \mu)$ will be called ε-optimal if $R(\pi, \lambda, \mu) \le R(x, \lambda, \mu) + \varepsilon$. For $\varepsilon = 0$ the above portfolio is called an optimal portfolio.

The following theorem (which is proved in Section 5) provides an existence result for the optimal portfolio.

THEOREM 2.1. *Let $F \in \mathcal{C}(M)$. Consider an American option with the continuous payoff process $Y(t) = F(t, S)$, $t \in [0, T]$. For any $\lambda > 0$, $0 < \mu < 1$ and $x \in \mathbb{R}_+$ there exists a portfolio strategy $\pi \in \mathcal{A}(x, \lambda, \mu)$ such that*

$$R(\pi, \lambda, \mu) = R(x, \lambda, \mu). \tag{2.12}$$

Next, we introduce the binomial models. Similar binomial models were used to approximate option prices and shortfall risks in the complete setup (see Kifer 2006; Dolinsky and Kifer 2008, 2011), i.e., in the absence of transaction costs.

For any n consider the n-step binomial market, which consists of a savings account $B^{(n)}(t)$ given by

$$B^{(n)}(t) \equiv B_0 > 0, \tag{2.13}$$

and of a risky stock $S^{(n)}$ given by the formulas $S^{(n)}(t) = S_0$ for $t \in [0, T/n)$ and

$$S^{(n)}(t) = S_0 \exp\left(\kappa \sqrt{\frac{T}{n}} \sum_{k=1}^{[nt/T]} \xi_k\right) \quad \text{if} \quad t \ge T/n, \tag{2.14}$$

where $\xi_1, \xi_2, \ldots$ are i.i.d. random variables taking values 1 and -1 with probabilities $p^{(n)} = (\exp((\kappa - \frac{2\vartheta}{\kappa})\sqrt{\frac{T}{n}}) + 1)^{-1}$ and $1 - p^{(n)} = (\exp((\frac{2\vartheta}{\kappa} - \kappa)\sqrt{\frac{T}{n}}) + 1)^{-1}$, respectively. Let $P_n = \{p^{(n)}, 1 - p^{(n)}\}^{\infty}$ be the corresponding product probability measure on the space of sequences $\Omega_\xi = \{-1, 1\}^{\infty}$. For any $k \ge 0$ let $\mathcal{F}^{\xi}_k = \sigma\{\xi_1, \ldots, \xi_k\}$, $(\mathcal{F}^{\xi}_0 = \{\emptyset, \Omega_\xi\})$. Denote by $\mathcal{T}_{0,n}$ the set of all stopping times with respect to the filtration $\mathcal{F}^{\xi}_k$ with values in $\{0, 1, \ldots, n\}$.

The n-step binomial market is active at the times $0, \frac{T}{n}, \frac{2T}{n}, \ldots, T$. As in the BS model, we assume that purchase (respectively, sale) of the risky asset is subject to a proportional transaction cost of rate λ (respectively, μ). Thus in the n-step binomial model a trading strategy with an initial capital $x \ge 0$ is a pair $\pi = (x, \{\gamma(k)\}^{n}_{k=0})$, where $\gamma(0) = 0$ and for any $k \ge 1$, $\gamma(k)$ is a random variable $\mathcal{F}^{\xi}_{k-1}$ measurable, which represents the number of stocks that the investor holds at time $\frac{kT}{n}$, before a transfer is made at this time. The portfolio value (in cash) of a trading strategy π is given by

$$V^{\pi}_{\lambda,\mu}(k) = x - (1+\lambda)\left(\gamma(k)^{-}S^{(n)}(kT/n) + \sum_{i=1}^{k}(\gamma(i)-\gamma(i-1))^{+}S^{(n)}((i-1)T/n)\right) \tag{2.15}$$
$$+ (1-\mu)\left(\gamma(k)^{+}S^{(n)}(kT/n) + \sum_{i=1}^{k}(\gamma(i)-\gamma(i-1))^{-}S^{(n)}((i-1)T/n)\right),$$
$$k = 0,\ 1, \dots, n.$$

Note that $V^{\pi}_{\lambda,\mu}(k)$ is the portfolio value at time $\frac{kT}{n}$, before a transfer is made at this time. A self-financing strategy π is called *admissible* if the following no-bankruptcy condition holds

$$V^{\pi}_{\lambda,\mu}(k) \geq 0 \quad \forall k \leq n. \tag{2.16}$$

The set of all *admissible* self-financing strategies with an initial capital x will be denoted by $\mathcal{A}^{(n)}(x,\lambda,\mu)$.

Consider an American contingent claim with the adapted payoff process

$$Y^{(n)}(k) = F\left(\frac{kT}{n}, S^{(n)}\right), \quad 0 \leq k \leq n. \tag{2.17}$$

For $\pi \in \mathcal{A}^{(n)}(x,\lambda,\mu)$ the shortfall risk is defined by

$$R_n(\pi,\lambda,u) = \max_{\tau\in\mathcal{T}_{0,n}} E_n\big[\big(Y^{(n)}(\tau) - V^{\pi}_{\lambda,\mu}(\tau)\big)^{+}\big], \tag{2.18}$$

where E_n is the expectation with respect to the probability measure P_n. The shortfall risk for an initial capital x is given by

$$R_n(x,\lambda,\mu) = \inf_{\pi\in\mathcal{A}^{(n)}(x,\lambda,\mu)} \max_{\tau\in\mathcal{T}_{0,n}} E_n\big[\big(Y^{(n)}(\tau) - V^{\pi}_{\lambda,\mu}(\tau)\big)^{+}\big]. \tag{2.19}$$

The following theorem is the main result of the paper and it states that the shortfall risk of an American option in the BS market with proportional transaction costs λ, μ can be approximated by a sequence of shortfall risks of an American options with the same proportional transaction costs in the binomial models defined earlier. This result has a practical value since for any n, the shortfall risk $R_n(x,\lambda,\mu)$ can be calculated by a dynamical programming algorithm, which is given in Section 3.

THEOREM 2.2. *For any* $\lambda > 0$, $0 < \mu < 1$ *and* $x \in \mathbb{R}_{+}$,

$$\lim_{n\to\infty} R_n(x,\lambda,\mu) = R(x,\lambda,\mu). \tag{2.20}$$

If $F \in \mathcal{C}_{Lip}(M)$ *then there exists constants* C_1, C_2, C_3 *(which do not depend on* x,λ,μ*), such that for any* $n \in \mathbb{N}$,

$$R_n(x,\lambda,\mu) - C_1 n^{-1/4}(\ln n)^{3/4} - \exp\left(\frac{C_2}{(\lambda+\mu)^2}\right) n^{-1/4} \leq R(x,\lambda,\mu) \tag{2.21}$$
$$\leq R_n(x,\lambda,\mu) + C_3 n^{-1/4}(\ln n)^{3/4}.$$

Next, we introduce a simple form of Skorohod embedding, which allows to consider the above-mentioned binomial markets and the BS model on the same probability space. Set $W^{*}(t) = \frac{\ln S(t)}{\kappa}$, $t \geq 0$, and for any $n \in \mathbb{N}$ define recursively $\theta^{(n)}_0 = 0$,

$\theta^{(n)}_{k+1} = \inf\{t > \theta^{(n)}_k : |W^*(t) - W^*(\theta^{(n)}_k)| = \sqrt{\frac{T}{n}}\}$. Observe (see Dolinsky and Kifer 2008) that for any k, $W^*(\theta^{(n)}_{k+1}) - W^*(\theta^{(n)}_k)$ is independent of $\mathcal{F}_{\theta^{(n)}_k}$ and excepts the values $\sqrt{\frac{T}{n}}$ and $-\sqrt{\frac{T}{n}}$, with probabilities $p^{(n)}$ and $1 - p^{(n)}$, respectively. For any n, define the map $\Pi_n : L^\infty(\mathcal{F}^\xi_n, P_n) \to L^\infty(\mathcal{F}_{\theta^{(n)}_n}, P)$ by $\Pi_n(U) = \tilde{U}$ so that if $U = f(\sqrt{\frac{T}{n}}\xi_1, \dots, \sqrt{\frac{T}{n}}\xi_n)$ for a function f on $\{\sqrt{\frac{T}{n}}, -\sqrt{\frac{T}{n}}\}^n$ then

$$\tilde{U} = f\big(W^*\big(\theta^{(n)}_1\big), W^*\big(\theta^{(n)}_2\big) - W^*\big(\theta^{(n)}_1\big), \dots, W^*\big(\theta^{(n)}_n\big) - W^*\big(\theta^{(n)}_{n-1}\big)\big).$$

Let $\mathcal{A}^{W,n}(x, \lambda, \mu)$ be the set of *admissible* self-financing strategies which managed on the set $\{0, \theta^{(n)}_1, \dots, \theta^{(n)}_n\}$ such that after the time $\theta^{(n)}_n$ the number of stocks in the portfolio is 0. Namely, $\pi = (x, \{\gamma(t)\}^\infty_{t=0}) \in \mathcal{A}^{W,n}(x, \lambda, \mu)$ if there exist random variables $u_1, \dots, u_n$ such that for any $i \geq 1$, u_i is $\mathcal{F}_{\theta^{(n)}_{i-1}}$ measurable and

$$\gamma(t) = \sum_{i=0}^{n-1} \mathbb{I}_{\theta^{(n)}_i < t \leq \theta^{(n)}_{i+1}} u_{i+1}, \tag{2.22}$$

where we set $\mathbb{I}_A = 1$ if an event A occurs and $\mathbb{I}_A = 0$ if not. We require that the corresponding wealth process, which is given by (2.8) satisfies the no-bankruptcy condition (2.9). The map Π_n allows us to define a function $\psi_n : \mathcal{A}^{(n)}(x, \lambda, \mu) \to \mathcal{A}^{W,n}(x, \lambda, \mu)$ which maps *admissible* self-financing strategies in the n-step binomial model to the set of *admissible* self-financing strategies in the BS model. Let $\pi = (x, \{\gamma(k)\}^n_{k=0}) \in \mathcal{A}^{(n)}(x, \lambda, \mu)$. Define $\psi_n(\pi) = (x, \{\tilde{\gamma}(t)\}^\infty_{t=0})$ by

$$\tilde{\gamma}(t) = \sum_{i=0}^{n-1} \mathbb{I}_{\theta^{(n)}_i < t \leq \theta^{(n)}_{i+1}} \Pi_n(\gamma(i+1)). \tag{2.23}$$

Let us show that $\tilde{\pi} := \psi_n(\pi)$ is an *admissible* portfolio. From (2.8), (2.15), and the equality $\Pi_n(S^{(n)}(kT/n)) = S(\theta^{(n)}_k)$, $k \leq n$ it follows that

$$V^{\tilde{\pi}}_{\lambda,\mu}\big(\theta^{(n)}_k\big) = \Pi_n\big(V^\pi_{\lambda,\mu}(k)\big) \geq 0, \quad k = 0, 1, \dots, n. \tag{2.24}$$

The portfolio strategy $\tilde{\pi}$ is managed only on the set $\{0, \theta^{(n)}_1, \dots, \theta^{(n)}_n\}$, and so it is clear that the wealth process $\{V^{\tilde{\pi}}_{\lambda,\mu}(t)\}^\infty_{t=0}$ is a supermartingale with respect to the measure $\tilde{P}$. Furthermore for any t, $V^{\tilde{\pi}}_{\lambda,\mu}(t) = V^{\tilde{\pi}}_{\lambda,\mu}(t \wedge \theta^{(n)}_n)$. This together with (2.24) gives

$$V^{\tilde{\pi}}_{\lambda,\mu}(t) \geq \tilde{E}\big(V^{\tilde{\pi}}_{\lambda,\mu}\big(\theta^{(n)}_n\big)\big| \mathcal{F}_{\theta^{(n)}_n \wedge t}\big) \geq 0. \tag{2.25}$$

Thus, $\psi_n(\pi)$ satisfies the no-bankruptcy condition, and so $\psi_n(\pi) \in \mathcal{A}^{W,n}(x, \lambda, \mu)$. If we restrict the portfolio $\psi_n(\pi)$ to the interval $[0, T]$, we get an element which belongs to $\mathcal{A}(x, \lambda, \mu)$. This restricted portfolio will be denoted by $\psi^T_n(\pi)$.

In Section 3 we prove that the optimal portfolios for the shortfall risk measure in the above-mentioned binomial models can be calculated by using a dynamical programming algorithm. The following result shows how to use these portfolios in order to construct "almost" optimal portfolios in the BS model.

THEOREM 2.3. *Let $\lambda > 0$, $0 < \mu < 1$ and $x \geq 0$. For any $n \in \mathbb{N}$, let $\pi_n = \pi_n(x, \lambda, \mu) \in \mathcal{A}^{(n)}(x, \lambda, \mu)$ be the optimal portfolio given by (3.19)–(3.20). Then*

$$\lim_{n\to\infty} R\big(\psi^T_n(\pi_n), \lambda, \mu\big) = R(x, \lambda, \mu). \tag{2.26}$$

If $F \in \mathcal{C}_{Lip}(M)$ *then there exists a constant* C_4 *such that for any* $n \in \mathbb{N}$,

$$(2.27)\quad R\big(\psi_n^T(\pi_n), \lambda, \mu\big) \le R(x, \lambda, \mu) + C_4 n^{-1/4}(\ln n)^{3/4} + \exp\left(\frac{C_2}{(\lambda+\mu)^2}\right) n^{-1/4}.$$

3. ANALYSIS OF THE BINOMIAL MODELS

In this section we provide a dynamical programming algorithm for the shortfall risks and the corresponding optimal portfolios in the binomial models. This dynamical programming algorithm will be essential for comparing the shortfall risks in the binomial models with the shortfall risk in the BS model. Throughout this section we assume that the transaction costs λ, μ are fixed.

Let $\pi = (x, \{\gamma(k)\}_{k=0}^n) \in \mathcal{A}^{(n)}(x, \lambda, \mu)$ for some $x \ge 0$ and $n \in \mathbb{N}$. From (2.15) it follows that

(3.1)
$$V^{\pi}_{\lambda,\mu}(k+1) = G\left(V^{\pi}_{\lambda,\mu}(k), \gamma(k)S^{(n)}(kT/n), (\gamma(k+1)-\gamma(k))S^{(n)}(kT/n), \exp\left(\kappa\sqrt{\frac{T}{n}}\xi_{k+1}\right)\right),$$
$$k = 0, 1, \dots, n-1$$

where

$$(3.2)\quad G(u, v, w, \rho) = u - (1-\mu)v^+ + (1+\lambda)v^- + (1-\mu)w^- - (1+\lambda)w^+ + \rho((1-\mu)(w+v)^+ - (1+\lambda)(w+v)^-).$$

For any $(u, v) \in \mathbb{R}_+ \times \mathbb{R}$, $0 < a < 1$ and $b > 0$ introduce the set $\mathcal{A}_{a,b}(u, v) = \{w \mid G(u, v, w, 1+b), G(u, v, w, 1-a) \ge 0\}$. By simple calculations we obtain that

(3.3)
$$\mathcal{A}_{a,b}(u, v) = \left[-v - \frac{u}{(1+\lambda)(1+b)-(1-\mu)}, \frac{(u-av(1-\mu))^+}{1+\lambda-(1-\mu)(1-a)} - \frac{(u-av(1-\mu))^-}{a(1-\mu)}\right]$$
if $v \ge 0$ and
$$\mathcal{A}_{a,b}(u, v) = \left[-\frac{(u+b(1+\lambda)v)^+}{(1+b)(1+\lambda)-(1-\mu)} + \frac{(u+b(1+\lambda)v)^-}{b(1+\lambda)}, -v + \frac{u}{1+\lambda-(1-\mu)(1-a)}\right]$$
if $v < 0$.

Set $a_n = 1 - \exp(-\kappa\sqrt{\frac{T}{n}})$ and $b_n = \exp(\kappa\sqrt{\frac{T}{n}}) - 1$. From (3.1) and the independence of ξ_{k+1} and $\mathcal{F}_k^\xi$ it follows that $\pi = (x, \{\gamma(k)\}_{k=0}^n) \in \mathcal{A}^{(n)}(x, \lambda, \mu)$ iff for any k, $\gamma(k)$ is $\mathcal{F}_{k-1}^\xi$ measurable ($\gamma(0) = 0$) and

$$(3.4)\quad (\gamma(k+1)-\gamma(k))S^{(n)}(kT/n) \in \mathcal{A}_{a_n,b_n}\left(V^{\pi}_{\lambda,\mu}(k), \gamma(k)S^{(n)}(kT/n)\right).$$

Next, we prove a technical lemma.

LEMMA 3.1. *Let* $0 < a, p < 1$, $b > 0$, *and* $H_1, H_2 : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ *be functions that satisfy the following conditions. For* $i = 1, 2$:

(i) H_i *is a continuous function.*
(ii) For any $v \in \mathbb{R}$, $H_i(\cdot, v)$ *is a nonincreasing function.*

(iii) H_i *is a piecewise linear function, which vanishes at infinity with respect to the first variable. Namely, there exist natural numbers* $N^{(i)}, M^{(i)} \in \mathbb{N}$ *and convex polyhedrals* $K_1^{(i)}, \dots, K_{N^{(i)}}^{(i)} \subset \mathbb{R}_+ \times \mathbb{R}$ *with pairwise disjoint interiors and* $\bigcup_{j=1}^{N^{(i)}} K_j^{(i)} = [0, M^{(i)}] \times \mathbb{R}$, *such that for any* $j \le N^{(i)}$

$$H_i(u, v) = c_j^{(i)} u + d_j^{(i)} v + e_j^{(i)} \quad \forall (u, v) \in K_j^{(i)}, \tag{3.5}$$

where $c_1^{(i)}, \dots, c_{N^{(i)}}^{(i)}, d_1^{(i)}, \dots, d_{N^{(i)}}^{(i)}, e_1^{(i)}, \dots, e_{N^{(i)}}^{(i)} \in \mathbb{R}$ *are constants.*

Define the function $H : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ *by*

$$\begin{aligned} H(u, v) = \inf_{w \in \mathcal{A}_{a,b}(u,v)} & \, pH_1(G(u, v, w, 1+b), (v+w)(1+b)) \\ & + (1-p)H_2\big(G(u, v, w, 1-a), (v+w)(1-a)\big). \end{aligned} \tag{3.6}$$

Then H *is satisfying the conditions (i)–(iii) above.*

Proof. Set $I(u, v, w) = pH_1(G(u, v, w, 1+b), (v+w)(1+b)) + (1-p)H_1(G(u, v, w, 1-a), (v+w)(1-a))$. Observe that $I(\cdot, u, v)$ is a nonincreasing function for any v, w. Clearly, for any $0 \le u_1 < u_2$ and $v \in \mathbb{R}$, $\mathcal{A}_{a,b}(u_1, v) \subseteq \mathcal{A}_{a,b}(u_2, v)$. Thus,

$$\begin{aligned} H(u_1, v) = \inf_{w \in \mathcal{A}_{a,b}(u_1,v)} I(u_1, v, w) &\ge \inf_{w \in \mathcal{A}_{a,b}(u_2,v)} I(u_1, v, w) \\ &\ge \inf_{w \in \mathcal{A}_{a,b}(u_2,v)} I(u_2, v, w) = H(u_2, v) \end{aligned} \tag{3.7}$$

and so, H satisfies condition (ii). Next, we prove continuity. Let $(u, v) \in \mathbb{R}_+ \times \mathbb{R}$ and $\{(u_n, v_n)\}_{n=1}^{\infty} \subset \mathbb{R}_+ \times \mathbb{R}$ such that $(u_n, v_n) \to (u, v)$ and $\lim_{n\to\infty} H(u_n, v_n)$ exists (possibly $\pm\infty$). For any n there exists (I is a continuous function) $w_n \in \mathcal{A}_{a,b}(u_n, v_n)$, which satisfies $I(u_n, v_n, w_n) = H(u_n, v_n)$. The sequence $\{w_n\}_{n=1}^{\infty}$ is bounded and hence, there exists a subsequence $\{w_{n_k}\}_{k=1}^{\infty}$, which converges to some w. From (3.3) it follows that $w \in \mathcal{A}_{a,b}(u, v)$ and so

$$H(u, v) \le I(u, v, w) = \lim_{n\to\infty} I(u_n, v_n, w_n) = \lim_{n\to\infty} H(u_n, v_n). \tag{3.8}$$

Choose $\tilde{w} \in \mathcal{A}_{a,b}(u, v)$ for which $I(u, v, \tilde{w}) = H(u, v)$. From (3.3) it follows that there exists a sequence $\tilde{w}_n \in \mathcal{A}_{a,b}(u_n, v_n)$, $n \in \mathbb{N}$ such that $\lim_{n\to\infty} \tilde{w}_n = \tilde{w}$. Thus,

$$H(u, v) = I(u, v, \tilde{w}) = \lim_{n\to\infty} I(u_n, v_n, \tilde{w}_n) \ge \lim_{n\to\infty} H(u_n, v_n). \tag{3.9}$$

From (3.8) and (3.9) we obtain that H is continuous. Finally, we prove that H satisfies condition (iii). For every $(u, v) \in \mathbb{R}_+ \times \mathbb{R}$ introduce the set

$$\begin{aligned} B(u, v) = &\left\{ w \in \mathcal{A}_{a,b}(u, v) | (G(u, v, w, 1+b), (v+w)(1+b)) \in \bigcup_{j=1}^{N^{(1)}} \partial K_j^{(1)} \right\} \bigcup \\ &\left\{ w \in \mathcal{A}_{a,b}(u, v) | (G(u, v, w, 1-a), (v+w)(1-a)) \in \bigcup_{j=1}^{N^{(2)}} \partial K_j^{(2)} \right\} \bigcup \partial \mathcal{A}_{a,b}(u, v). \end{aligned}$$

Fix u, v, and let $B(u, v) = \{w_1, \dots, w_k\}$ for some $k \in \mathbb{N}$ and $w_1 < w_2 < \cdots < w_k$. From (3.5) it follows that for any $i < k$, the function $I(u, v, \cdot)$ is linear on the interval $[w_i, w_{i+1}]$, and so

(3.10) $$H(u,v)=\min_{w\in B(u,v)} I(u,v,w).$$

Note that there exists a finite sequence of real numbers $\alpha_1,\dots,\alpha_N,\beta_1,\dots,\beta_N,\delta_1,\dots,\delta_N$ such that for any (u,v), $B(u,v)\subseteq\{\alpha_j u+\beta_j v+\delta_j\,|\,j\le N\}$. This together with (3.10) gives that there exists a finite sequence of real numbers $\Phi_1,\dots,\Phi_m$, $\Delta_1,\dots,\Delta_m$,$\Theta_1,\dots,\Theta_m$ such that for any $(u,v)\in\mathbb{R}_+\times\mathbb{R}$

(3.11) $$H(u,v)=\Phi_j u+\Delta_j v+\Theta_j$$

for some j, which depends on u,v. From (3.3), $-v\in\mathcal{A}_{a,b}(u,v)$ and so

(3.12) $$H(u,v)\le I(u,v,-v)=pH_1(u,0)+(1-p)H_2(u,0)\le\max(H_1(u,0),H_2(u,0)).$$

From (3.11) and (3.12) and the fact that H is continuous we conclude that H satisfies condition (iii). This completes the proof. □

Next, we fix n and consider the n-step binomial model. For any $\pi\in\mathcal{A}^{(n)}(x,\lambda,\mu)$ define a sequence of random variables $\{U^\pi(k)\}_{k=0}^n$ by

(3.13)
$$U^\pi(n)=\left(Y^{(n)}(n)-V^\pi_{\lambda,\mu}(n)\right)^+,\quad\text{and}\quad\text{for } k<n$$
$$U^\pi(k)=\max\left(E_n\left(U^\pi(k+1)\middle|\,\mathcal{F}^\xi_k\right),\left(Y^{(n)}(k)-V^\pi_{\lambda,\mu}(k)\right)^+\right).$$

Applying standard results for optimal stopping (see Peskir and Shiryaev 2006) for the process $(Y^{(n)}(k)-V^\pi_{\lambda,\mu}(k))^+$, $k=0,1,\dots,n$ we obtain

(3.14) $$U^\pi(0)=\max_{\tau\in\mathcal{T}_{0,n}} E_n\left[\left(Y^{(n)}(\tau)-V^\pi_{\lambda,\mu}(\tau)\right)^+\right]=R_n(\pi,\lambda,\mu).$$

For any $0\le k\le n$ let $\phi^{(n)}_k:\{-1,1\}^k\to\mathbb{R}_+$ such that

(3.15) $$\phi^{(n)}_k(\xi_1,\dots,\xi_k)=Y^{(n)}(k).$$

Define a sequence of functions $J^{(n)}_k:\mathbb{R}_+\times\mathbb{R}\times\{-1,1\}^k\to\mathbb{R}_+$, $k=0,1,\dots,n$ by the following backward relations. For any $z_1,\dots,z_n\in\{-1,1\}$ and $(u,v)\in\mathbb{R}_+\times\mathbb{R}$

(3.16) $$J^{(n)}_n(u,v,z_1,\dots,z_n)=\left(\phi^{(n)}_n(z_1,\dots,z_n)-u\right)^+$$

and for $k<n$

(3.17)
$$J^{(n)}_k(u,v,z_1,\dots,z_k)=\max\Bigg(\left(\phi^{(n)}_k(z_1,\dots,z_k)-u\right)^+,$$
$$\inf_{w\in\mathcal{A}_{a_n,b_n}(u,v)}\Bigg[p^{(n)}J^{(n)}_{k+1}\left(G(u,v,w,1+b_n),(1+b_n)(v+w),z_1,\dots,z_k,1\right)$$
$$+(1-p^{(n)})J^{(n)}_{k+1}\left(G(u,v,w,1-a_n),(1-a_n)(v+w),z_1,\dots,z_k,-1\right)\Bigg]\Bigg),$$

where $p^{(n)}$ was defined after (2.14). From Lemma 3.1 it follows (by backward induction) that for any $k\le n$ and $z_1,\dots,z_k\in\{-1,1\}$ the function $H(\cdot,\cdot\cdot):=J^{(n)}_k(\cdot,\cdot\cdot,z_1,\dots,z_k)$ is satisfying conditions (i)–(iii), which were introduced in Lemma 3.1. In particular it is continuous. This fact allows us to define the functions $h_k^{(n)} : \mathbb{R}_+ \times \mathbb{R} \times \{-1, 1\}^k \to \mathbb{R}$, $0 \le k < n$ by

(3.18)
$$\begin{aligned} h_k^{(n)}(u, v, z_1, \ldots, z_k) &= \operatorname*{argmin}_{w \in \mathcal{A}_{a_n, b_n}(u,v)} \big[p^{(n)} J_{k+1}^{(n)}(G(u, v, w, 1 + b_n), (1 + b_n)(v + w), z_1, \ldots, z_k, 1) \\ &\qquad + (1 - p^{(n)}) J_{k+1}^{(n)}(G(u, v, w, 1 - a_n), (1 - a_n)(v + w), z_1, \ldots, z_k, -1)\big]. \end{aligned}$$

Let $x \ge 0$ be an initial capital. Define $\pi = \pi_n(x, \lambda, \mu) = (x, \{\gamma(k)\}_{k=0}^n)$ by

$$\gamma(0) = 0, \quad V_{\lambda,\mu}^{\pi}(0) = x, \tag{3.19}$$

and for $k < n$,

(3.20)
$$\begin{aligned} \gamma(k+1) &= \gamma(k) + \frac{1}{S^{(n)}(kT/n)} h_k^{(n)}\left(V_{\lambda,\mu}^{\pi}(k), \gamma(k)S^{(n)}(kT/n), \xi_1, \ldots, \xi_k\right), \\ V_{\lambda,\mu}^{\pi}(k+1) &= G\left(V_{\lambda,\mu}^{\pi}(k), \gamma(k)S^{(n)}(kT/n), (\gamma(k+1) - \gamma(k))S^{(n)}(kT/n), \exp\left(\kappa\sqrt{\frac{T}{n}}\xi_{k+1}\right)\right). \end{aligned}$$

PROPOSITION 3.2. *For any $n \in \mathbb{N}$ and $x \ge 0$*

$$R_n(\pi_n(x, \lambda, \mu), \lambda, \mu) = R_n(x, \lambda, \mu) = J_0^{(n)}(x, 0). \tag{3.21}$$

Proof. Fix $n \in \mathbb{N}$ and $x \ge 0$. Set $\pi = \pi_n(x, \lambda, \mu) = (x, \gamma)$ and let $\tilde{\pi} = (x, \tilde{\gamma}) \in \mathcal{A}^{(n)}(x, \lambda, \mu)$ be an arbitrary portfolio. First we prove by backward induction that for any $0 \le k \le n$,

$$J_k^{(n)}\left(V_{\lambda,\mu}^{\pi}(k), \gamma(k)S^{(n)}(kT/n), \xi_1, \ldots, \xi_k\right) = U^{\pi}(k) \tag{3.22}$$

and

$$J_k^{(n)}\big(V_{\lambda,\mu}^{\tilde{\pi}}(k), \tilde{\gamma}(k)S^{(n)}(kT/n), \xi_1, \ldots, \xi_k\big) \le U^{\tilde{\pi}}(k). \tag{3.23}$$

For $k = n$, we obtain from (3.13), (3.15), and (3.16) that the relations (3.22) and (3.23) hold with equality. Suppose that (3.22) and (3.23) hold true for $k + 1$ and let us prove them for k. Set,

$$\begin{aligned} &\Upsilon = \gamma(k)S^{(n)}(kT/n), \quad \tilde{\Upsilon} = \tilde{\gamma}(k)S^{(n)}(kT/n), \\ &\Gamma = h_k^{(n)}\big(V_{\lambda,\mu}^{\pi}(k), \Upsilon, \xi_1, \ldots, \xi_k\big) \quad \text{and} \quad \tilde{\Gamma} = (\tilde{\gamma}(k+1) - \tilde{\gamma}(k))S^{(n)}(kT/n). \end{aligned}$$

From (3.18)–(3.20) and the induction assumption it follows that

(3.24)
$$
\begin{aligned}
E_n\big(U^{\pi}(k+1)\big|\mathcal{F}^{\xi}_k\big) &= E_n\big(J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\Upsilon,\Gamma,\exp(\kappa\sqrt{T/n}\xi_{k+1})\big),\\
&\qquad \times(\Gamma+\Upsilon)\exp(\kappa\sqrt{T/n}\xi_{k+1}),\xi_1,\dots,\xi_{k+1}\big)\big|\mathcal{F}^{\xi}_k\big)\\
&= p^{(n)}J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\Upsilon,\Gamma,1+b_n\big),(\Gamma+\Upsilon)(1+b_n),\xi_1,\dots,\xi_k,1\big)\\
&\quad +(1-p^{(n)})J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\Upsilon,\Gamma,1-a_n\big),(\Gamma+\Upsilon)(1-a_n),\xi_1,\dots,\xi_k,-1\big)\\
&= \min_{w\in\mathcal{A}_{a_n,b_n}(V^{\pi}_{\lambda,\mu}(k),\Upsilon)}\big[p^{(n)}J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\Upsilon,w,1+b_n\big),(w+\Upsilon)(1+b_n),\xi_1,\dots,\xi_k,1\big)\\
&\qquad +(1-p^{(n)})J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\Upsilon,w,1-a_n\big),\\
&\qquad (w+\Upsilon)(1-a_n),\xi_1,\dots,\xi_k,-1\big)\big],
\end{aligned}
$$

where the last equality follows from (3.18) and (3.20). From (3.4) it follows that $\tilde{\Gamma}\in\mathcal{A}_{a_n,b_n}(V^{\tilde{\pi}}_{\lambda,\mu}(k),\tilde{\Upsilon})$, and so from the induction assumption

(3.25)
$$
\begin{aligned}
E_n\big(U^{\tilde{\pi}}(k+1)\big|\mathcal{F}^{\xi}_k\big) &= E_n\big(J^{(n)}_{k+1}\big(G\big(V^{\tilde{\pi}}_{\lambda,\mu}(k),\tilde{\Upsilon},\tilde{\Gamma},\exp(\kappa\sqrt{T/n}\xi_{k+1})\big),\\
&\qquad (\tilde{\Gamma}+\tilde{\Upsilon})\exp(\kappa\sqrt{T/n}\xi_{k+1}),\xi_1,\dots,\xi_{k+1}\big)\big|\mathcal{F}^{\xi}_k\big)\\
&= p^{(n)}J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\tilde{\Upsilon},\tilde{\Gamma},1+b_n\big),(\tilde{\Gamma}+\tilde{\Upsilon})(1+b_n),\xi_1,\dots,\xi_k,1\big)\\
&\quad +(1-p^{(n)})J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\tilde{\Upsilon},\tilde{\Gamma},1-a_n\big),(\tilde{\Gamma}+\tilde{\Upsilon})(1-a_n),\xi_1,\dots,\xi_k,-1\big)\\
&\geq \min_{w\in\mathcal{A}_{a_n,b_n}(V^{\pi}_{\lambda,\mu}(k),\tilde{\Upsilon})}\big[p^{(n)}J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\tilde{\Upsilon},w,1+b_n\big),\\
&\qquad (w+\tilde{\Upsilon})(1+b_n),\xi_1,\dots,\xi_k,1\big)\\
&\qquad +(1-p^{(n)})J^{(n)}_{k+1}\big(G\big(V^{\pi}_{\lambda,\mu}(k),\tilde{\Upsilon},w,1-a_n\big),\\
&\qquad (w+\tilde{\Upsilon})(1-a_n),\xi_1,\dots,\xi_k,-1\big)\big].
\end{aligned}
$$

Combining (3.13), (3.15)–(3.17), and (3.24) and (3.25) we obtain that (3.22) and (3.23) hold true. Next, by using (3.22) and (3.23) for $k=0$ and (3.14) it follows that for any $\tilde{\pi}\in\mathcal{A}^{(n)}(x,\lambda,\mu)$

$$R_n(\pi,\lambda,\mu)=U^{\pi}(0)=J^{(n)}_0(x,0)\leq U^{\tilde{\pi}}(0)=R_n(\tilde{\pi},\lambda,\mu).$$

Thus $R_n(x,\lambda,\mu)=R_n(\pi,\lambda,\mu)=J^{(n)}_0(x,0)$, as required. □

COROLLARY 3.3. *From Lemma 3.1 and Proposition 3.2 we obtain that the function $R_n(\cdot,\lambda,\mu)=J^{(n)}_0(\cdot,0)$ is a continuous nonincreasing piecewise linear function vanishing at ∞. Namely, there exists $N\in\mathbb{N}$, $c_1,\dots,c_N\leq 0$, $d_1,\dots,d_N\in\mathbb{R}$ and $0=\beta_1<\beta_2<\cdots<\beta_{N+1}<\infty$ such that for any $x\in\mathbb{R}_+$, $R_n(x,\lambda,\mu)=\sum_{i=1}^N\mathbb{I}_{[\beta_i,\beta_{i+1})}(c_ix+d_i)$.*

4. PROOF OF LIMIT THEOREMS

In this section we complete the proof of Theorems 2.2 and 2.3. We start with the following lemma, which provides a bound for $\tilde{E}(\sup_{0\leq t\leq T}(\int_0^t\gamma(u)\,dS(u))^2)$ for an *admissible* portfolio strategy $\pi = (x, \gamma)$. Recall that $\tilde{E}$ is the expectation with respect to the unique martingale measure $\tilde{P}$ of the complete BS model. A similar result (without explicit bounds) was proved in Cvitanic and Karatzas (1996). Since our setup is a bit different, we give a self-contained proof, which follows the lines of the original proof in Cvitanic and Karatzas (1996).

LEMMA 4.1. *For any* $\lambda > 0$, $0 < \mu < 1$, $x \in \mathbb{R}_+$, *and* $\pi = (x, \gamma) \in \mathcal{A}(x, \lambda, \mu)$,

$$\tilde{E}\left(\sup_{0\le t\le T}\left(\int_0^t \gamma(u)\,dS(u)\right)^2\right) \le 8\left(\frac{\kappa x}{\lambda+\mu}\right)^2 T\exp\left(2\kappa^2 T\left(\frac{1+\lambda}{\mu+\lambda}\right)^2\right). \tag{4.1}$$

Proof. Fix λ, μ, x, and $\pi = (x, \gamma) \in \mathcal{A}(x, \lambda, \mu)$. From (2.7)–(2.9)

$$\begin{aligned} &x - (1+\lambda)\int_0^t S(u)\,d\gamma(u) - (\mu+\lambda)\int_0^t S(u)\,d\gamma^-(u) + (1+\lambda) \\ &\quad\times \gamma(t)S(t) \ge 0 \ \text{ and } \ x - (1-\mu)\int_0^t S(u)\,d\gamma(u) - (\mu+\lambda) \\ &\quad\times \int_0^t S(u)\,d\gamma^+(u) + (1-\mu)\gamma(t)S(t) \ge 0, \quad \forall t \in [0, T]. \end{aligned} \tag{4.2}$$

From the integration by parts formula we get that for any $t \in [0, T]$, $\gamma(t)S(t) = \int_0^t S(u)\,d\gamma(u) + \int_0^t \gamma(u)\,dS(u)$. This together with (4.2) yields

$$\begin{aligned} &\int_0^t S(u)\,d\gamma^-(u) \le \frac{1}{\lambda+\mu}\left(x + (1+\lambda)\int_0^t \gamma(u)\,dS(u)\right) \quad \text{and} \\ &\int_0^t S(u)\,d\gamma^+(u) \le \frac{1}{\lambda+\mu}\left(x + (1-\mu)\int_0^t \gamma(u)\,dS(u)\right), \quad \forall t \in [0, T]. \end{aligned} \tag{4.3}$$

Consequently,

$$\begin{aligned} \gamma(t)S(t) &\ge \int_0^t \gamma(u)\,dS(u) - \int_0^t S(u)\,d\gamma^-(u) \ge -\frac{x}{\lambda+\mu} - \frac{1-\mu}{\mu+\lambda} \\ &\quad\times \int_0^t \gamma(u)\,dS(u) \ \text{ and } \ \gamma(t)S(t) \le \int_0^t \gamma(u)\,dS(u) + \int_0^t S(u)\,d\gamma^+(u) \\ &\le \frac{x}{\lambda+\mu} + \frac{1+\lambda}{\mu+\lambda}\int_0^t \gamma(u)\,dS(u), \quad \forall t \in [0, T], \end{aligned} \tag{4.4}$$

which yields

$$|\gamma(t)S(t)|^2 \le 2\left(\frac{x}{\lambda+\mu}\right)^2 + 2\left(\frac{1+\lambda}{\mu+\lambda}\right)^2\left(\int_0^t \gamma(u)\,dS(u)\right)^2. \tag{4.5}$$

For any $n \in \mathbb{N}$, define the stopping time $\tau_n = \inf\{t \mid \gamma(t)S(t) \ge n\} \wedge T$. Set $\alpha_n(t) = \tilde{E}\big(\int_0^t \gamma(u)\mathbb{I}_{u\le\tau_n}\,dS(u)\big)^2$, $t \in [0, T]$. From (4.5) and the Itô isometry we obtain

$$
\begin{aligned}
(4.6)\quad \alpha_n(t) &= \kappa^2 \int_0^t \tilde{E}((\gamma(u)S(u))^2 \mathbb{I}_{u\le\tau_n})\,du \le \kappa^2 \int_0^t \tilde{E}((\gamma(u\wedge\tau_n)S(u\wedge\tau_n))^2)\,du \\
&\le 2\kappa^2\left(\frac{x}{\lambda+\mu}\right)^2 T + 2\kappa^2\left(\frac{1+\lambda}{\mu+\lambda}\right)^2 \int_0^t \tilde{E}\left(\left(\int_0^{u\wedge\tau_n}\gamma(v)\,dS(v)\right)^2\right)du \\
&= 2\left(\frac{\kappa x}{\lambda+\mu}\right)^2 T + 2\kappa^2\left(\frac{1+\lambda}{\mu+\lambda}\right)^2 \int_0^t \alpha_n(u)\,du.
\end{aligned}
$$

From Gronwall's inequality and (4.6) we get $\alpha_n(T) \le 2(\frac{\kappa x}{\lambda+\mu})^2 T \exp(2\kappa^2 T(\frac{1+\lambda}{\mu+\lambda})^2)$. Note that $\lim_{n\to\infty}\tau_n = T$, and so $\lim_{n\to\infty}\gamma(t)S(t)\mathbb{I}_{t\le\tau_n} = \gamma(t)S(t)$ a.s. in $dt d\tilde{P}$. Thus from Fubini's theorem and (4.6), $\tilde{E}\int_0^T(\gamma(u)S(u))^2 du < \infty$. We conclude that the local martingale $\{\int_0^t \gamma(u)\,dS(u)\}_{t=0}^T$ is a square integrable martingale. From Doob's inequality and Fatou's lemma

$$
\begin{aligned}
&\tilde{E}\left(\sup_{0\le t\le T}\left(\int_0^t \gamma(u)\,dS(u)\right)^2\right) \le 4\tilde{E}\left(\left(\int_0^T \gamma(u)\,dS(u)\right)^2\right) \\
&\quad = 4\tilde{E}\left(\lim_{n\to\infty}\left(\int_0^{T\wedge\tau_n}\gamma(u)\,dS(u)\right)^2\right) = 4\tilde{E}\left(\lim_{n\to\infty}\left(\int_0^T \gamma(u)\mathbb{I}_{u\le\tau_n}\,dS(u)\right)^2\right) \\
&\quad \le 4\liminf_{n\to\infty}\alpha_n(T) \le 8\left(\frac{\kappa x}{\lambda+\mu}\right)^2 T\exp\left(2\kappa^2 T\left(\frac{1+\lambda}{\mu+\lambda}\right)^2\right).
\end{aligned}
$$

□

Next, fix $\lambda > 0$ and $0 < \mu < 1$. Set $\lambda_n = (1+\lambda)\exp(2\kappa\sqrt{\frac{T}{n}}) - 1$ and $\mu_n = 1-(1-\mu)\exp(-2\kappa\sqrt{\frac{T}{n}})$, $n\in\mathbb{N}$. Since $\lambda_n \ge \lambda$ and $\mu_n \ge \mu$ we have $R(x,\lambda_n,\mu_n) \ge R(x,\lambda,\mu)$, $x\in\mathbb{R}_+$. The following result provides an estimate from above for the term $R(x,\lambda_n,\mu_n) - R(x,\lambda,\mu)$.

LEMMA 4.2. *There exists a constant $\tilde{C}_1$ such that for any $n\in\mathbb{N}$ and $x\ge 0$,*

$$
(4.7)\qquad R(x,\lambda_n,\mu_n) - R(x,\lambda,\mu) \le \tilde{C}_1 n^{-1/4}\left(1+\exp\left(\frac{\tilde{C}_1}{(\mu+\lambda)^2}\right)\right).
$$

Proof. If $x = 0$ then $R(x,\lambda_n,\mu_n) = R(x,\lambda,\mu) = \sup_{\tau\in\mathcal{T}_{[0,T]}} EY(\tau)$, and so the statement is trivial. Fix an initial capital $x > 0$ and $n\in\mathbb{N}$. Choose $\delta > 0$. There exists a portfolio $\pi = (x,\gamma)\in\mathcal{A}(x,\lambda,\mu)$ such that

$$
(4.8)\qquad R(\pi,\lambda,\mu) < R(x,\lambda,\mu) + \delta.
$$

Define the stopping times

$$
(4.9)\qquad
\begin{aligned}
\sigma_1 &= \inf\left\{t \,\middle|\, \int_0^t S(u)\,d\gamma^-(u) + \gamma(t)^+ S(t) \ge \frac{xn^{-1/4}}{2(\mu_n-\mu)}\right\}\wedge T \quad\text{and} \\
\sigma_2 &= \inf\left\{t \,\middle|\, \int_0^t S(u)\,d\gamma^+(u) + \gamma(t)^- S(t) \ge \frac{xn^{-1/4}}{2(\lambda_n-\lambda)}\right\}\wedge T.
\end{aligned}
$$

The stochastic processes $\{\int_0^t S(u)\,d\gamma^-(u) + \gamma(t)^+ S(t)\}_{t=0}^T$ and $\{\int_0^t S(u)\,d\gamma^+(u) + \gamma(t)^- S(t)\}_{t=0}^T$ are left continuous, and so for any $t \le T$,

$$\begin{aligned} &\int_0^{t\wedge\sigma} S(u)\,d\gamma^-(u) + \gamma(t)^+ S(t) \le \frac{xn^{-1/4}}{2(\mu_n - \mu)} \quad \text{and} \\ &\int_0^{t\wedge\sigma} S(u)\,d\gamma^+(u) + \gamma(t)^- S(t) \le \frac{xn^{-1/4}}{2(\lambda_n - \lambda)}, \end{aligned} \tag{4.10}$$

where $\sigma = \sigma_1 \wedge \sigma_2$. Consider the portfolio $\tilde{\pi} = (x, \{(1 - n^{-1/4})\gamma(t)\mathbb{I}_{t\le\sigma}\}_{t=0}^T)$. From (2.8) and (4.10) we obtain that for any $t \in [0, T]$,

$$\begin{aligned} V^{\tilde{\pi}}_{\lambda_n,\mu_n}(t) = {} & V^{\tilde{\pi}}_{\lambda_n,\mu_n}(t\wedge\sigma) = (1 - n^{-1/4}) V^{\pi}_{\lambda,\mu}(t\wedge\sigma) + xn^{-1/4} \\ & - (\mu_n - \mu)\left(\int_0^{t\wedge\sigma} S(u)\,d\gamma^-(u) + \gamma(t)^+ S(t)\right) - (\lambda_n - \lambda) \\ & \times \left(\int_0^{t\wedge\sigma} S(u)\,d\gamma^+(u) + \gamma(t)^- S(t)\right) \ge (1 - n^{-1/4}) V^{\pi}_{\lambda,\mu}(t\wedge\sigma). \end{aligned} \tag{4.11}$$

Thus $\tilde{\pi} \in \mathcal{A}(x, \lambda_n, \mu_n)$. From (2.3), (4.8), and (4.11)

$$\begin{aligned} R(x, \lambda_n, \mu_n) & \le \sup_{\tau\in\mathcal{T}_{[0,T]}} E\big[\big(Y(\tau) - (1 - n^{-1/4}) V^{\pi}_{\lambda,\mu}(\tau\wedge\sigma)\big)^+\big] \\ & \le \sup_{\tau\in\mathcal{T}_{[0,T]}} E\big[\big(Y(\tau) - V^{\pi}_{\lambda,\mu}(\tau\wedge\sigma)\big)^+\big] + n^{-1/4} E \sup_{0\le t\le T} Y(t) \\ & \le R(x, \lambda, \mu) + \delta + \tilde{E}\left((\mathbb{I}_{\sigma<T} + n^{-1/4}) Z^{-1}(T) \sup_{0\le t\le T} Y(t)\right). \end{aligned} \tag{4.12}$$

From (4.12) and the Cauchy–Schwarz inequality we obtain that there exists a constant $\hat{C}_1$ such that

$$R(x, \lambda_n, \mu_n) - R(x, \lambda, \mu) \le \delta + \hat{C}_1\left(\sqrt{\tilde{P}(\sigma < T)} + n^{-1/4}\right). \tag{4.13}$$

Set

$$\begin{aligned} \Gamma_1 &= \sup_{0\le t\le T}\left(\int_0^t S(u)\,d\gamma^-(u) + \gamma(t)^+ S(t)\right) \quad \text{and} \\ \Gamma_2 &= \sup_{0\le t\le T}\left(\int_0^t S(u)\,d\gamma^+(u) + \gamma(t)^- S(t)\right). \end{aligned}$$

From (4.3) and (4.4) we obtain

$$\begin{aligned} \Gamma_1 &\le \frac{2}{\lambda+\mu}\left(x + (1+\lambda)\sup_{0\le t\le T}\left|\int_0^t \gamma(u)\,dS(u)\right|\right) \quad \text{and} \\ \Gamma_2 &\le \frac{2}{\lambda+\mu}\left(x + (1-\mu)\sup_{0\le t\le T}\left|\int_0^t \gamma(u)\,dS(u)\right|\right). \end{aligned} \tag{4.14}$$

There exists a constant $\hat{C}_2$ such that $(\mu_n - \mu) \le \hat{C}_2 \frac{1-\mu}{\sqrt{n}}$ and $(\lambda_n - \lambda) \le \hat{C}_2 \frac{1+\lambda}{\sqrt{n}}$. This together with Lemma 4.1, Chebyshev's Inequality, and (4.14) implies

(4.15)
$$\begin{aligned}
\tilde{P}(\sigma < T) &\le \tilde{P}\left(\Gamma_1 \ge \frac{xn^{-1/4}}{2(\mu_n - \mu)}\right) + \tilde{P}\left(\Gamma_2 \ge \frac{xn^{-1/4}}{2(\lambda_n - \lambda)}\right) \\
&\le \frac{4\hat{C}_2^2(1-\mu)^2}{\sqrt{n}x^2}\left(\frac{8x^2}{(\lambda+\mu)^2} + \frac{64\kappa^2 x^2(1+\lambda)^2}{(\lambda+\mu)^4} T \exp\left(2\kappa^2 T\left(\frac{1+\lambda}{\mu+\lambda}\right)^2\right)\right) \\
&\quad + \frac{4\hat{C}_2^2(1+\lambda)^2}{\sqrt{n}x^2}\left(\frac{8x^2}{(\lambda+\mu)^2} + \frac{64\kappa^2 x^2(1-\mu)^2}{(\lambda+\mu)^4} T \exp\left(2\kappa^2 T\left(\frac{1+\lambda}{\mu+\lambda}\right)^2\right)\right) \\
&\le \hat{C}_3 n^{-1/2}\left(1 + \exp\left(\frac{\hat{C}_3}{(\mu+\lambda)^2}\right)\right),
\end{aligned}$$

for some constant $\hat{C}_3$. Since $\delta > 0$ is arbitrary then by combining (4.13) and (4.15) we conclude the proof. $\square$

The next step is to compare the shortfall risk in the BS model with the shortfall risks in the binomial models. We start with some technical preparation. For any $n \in \mathbb{N}$ and $0 \le k \le n$ introduce the finite σ-algebra $\mathcal{G}_k^n = \sigma\{W^*(\theta_1^{(n)}), \dots, W^*(\theta_k^{(n)})\}$ with $\mathcal{G}_0^n = \{\emptyset, \Omega_W\}$ being the trivial σ-algebra. Let $\mathcal{S}_{0,n}$ and $\mathcal{T}_{0,n}^W$ be the sets of all stopping times with values in the set $\{0, 1, \dots, n\}$ with respect to the filtrations $\{\mathcal{G}_k^n\}_{k=0}^n$ and $\{\mathcal{F}_{\theta_k^{(n)}}\}_{k=0}^n$, respectively. For any $n \in \mathbb{N}$, set

$$S^{W,n}(t) = S(\theta_k^{(n)}), \quad kT/n \le t < (k+1)T/n, \quad k = 0, 1, \dots, n. \tag{4.16}$$

Define

$$Y^{W,n}(t) = F(t, S^{W,n}), \quad t \in [0, T]. \tag{4.17}$$

Note that for any $0 \le k \le n$

$$\begin{aligned}
Y^{W,n}(kT/n) = \phi_k^{(n)}\Bigg(&\sqrt{\frac{n}{T}} W^*(\theta_1^{(n)}), \sqrt{\frac{n}{T}}\left(W^*(\theta_2^{(n)}) - W^*(\theta_1^{(n)})\right), \dots, \\
&\sqrt{\frac{n}{T}}\left(W^*(\theta_k^{(n)}) - W^*(\theta_{k-1}^{(n)})\right)\Bigg),
\end{aligned} \tag{4.18}$$

where the function $\phi_k^{(n)}$ was defined in (3.15). We will need the following estimates, which were obtained in Kifer (2006) (see (4.7), (4.8), and (4.25) and lemmas 3.2, 3.3 therein). For any $n \in \mathbb{N}$ and $a \in \mathbb{R}$,

$$\begin{aligned}
&Ee^{|a|(\theta_n^{(n)} \vee T)} \le e^{|a|\hat{C}_4 T}, \quad E \sup_{0 \le t \le \theta_n^{(n)} \vee T} \exp(aW(t)) \le 2e^{a^2 \hat{C}_5 T}, \\
&E\left(\max_{0 \le k \le n}\left|\theta_k^{(n)} - \frac{kT}{n}\right|^2\right) \le \frac{\hat{C}_6}{n}, \quad \text{and if} \quad F \in \mathcal{C}_{Lip}(M) \\
&\sup_{\tau \in \mathcal{T}_{0,n}^W} E\left|Y^{W,n}(\tau T/n) - Y(\theta_\tau^{(n)})\right| \le \hat{C}_7 n^{-1/4}(\ln n)^{3/4}
\end{aligned} \tag{4.19}$$

for some constants $\hat{C}_4$, $\hat{C}_5$, $\hat{C}_6$, and $\hat{C}_7$. In particular the sequence $\{\max_{0\le k\le n} |\theta_k^{(n)} - \frac{kT}{n}|\}_{n=1}^{\infty}$ converges to 0 in probability, and so

$$\lim_{n\to\infty} \sup_{0\le t\le T} |S^{W,n}(t) - S(t)| = 0 \text{ in probability.} \tag{4.20}$$

Recall the set $\mathcal{A}^{W,n}(x, \lambda, \mu)$ which was introduced before equation (2.22). Define

$$R^{W,n}(x, \lambda, \mu) = \inf_{\pi\in\mathcal{A}^{W,n}(x,\lambda,\mu)} \sup_{\tau\in\mathcal{T}_{0,n}^W} E\big[\big(Y^{W,n}(\tau T/n) - V_{\lambda,\mu}^{\pi}\big(\theta_\tau^{(n)}\big)\big)^+\big]. \tag{4.21}$$

From (2.8) it follows that for any $\pi = (x, \{\gamma(t)\}_{t=0}^{\infty}) \in \mathcal{A}^{W,n}(x, \lambda, \mu)$,

$$V_{\lambda,\mu}^{\pi}(\theta_{k+1}^{(n)}) = G\big(V_{\lambda,\mu}^{\pi}(k), \gamma\big(\theta_k^{(n)}\big)S\big(\theta_k^{(n)}\big), \Upsilon, \exp\big(\kappa\big(W^*\big(\theta_{k+1}^{(n)}\big) - W^*\big(\theta_k^{(n)}\big)\big)\big)\big), \tag{4.22}$$

where $\Upsilon = (\gamma(\theta_{k+1}^{(n)}) - \gamma(\theta_k^{(n)}))S(\theta_k^{(n)})$ and G was introduced in (3.2). Combining similar arguments to those of Section 3 (replace $\{\xi_i\}_{i=1}^n$, $\{S^{(n)}(\frac{iT}{n})\}_{i=0}^n$, and $\{\mathcal{F}_i^{\xi}\}_{i=0}^n$ by $\{\sqrt{\frac{n}{T}}(W^*(\theta_i^{(n)}) - W^*(\theta_{i-1}^{(n)}))\}_{i=1}^n$, $\{S(\theta_i^{(n)})\}_{i=0}^n$, and $\{\mathcal{F}_{\theta_i^{(n)}}\}_{i=0}^n$, respectively) with (4.18), (4.22), and the independence of $W^*(\theta_{k+1}^{(n)}) - W^*(\theta_k^{(n)})$ and $\mathcal{F}_{\theta_k^{(n)}}$, we obtain

$$R^{W,n}(x, \lambda, \mu) = J_0^{(n)}(x, 0) = R_n(x, \lambda, \mu) \quad \forall x \ge 0. \tag{4.23}$$

The above equality is essential for proving the following result.

LEMMA 4.3. *For any initial capital* $x \ge 0$,

$$\limsup_{n\to\infty} R_n(x, \lambda, \mu) - R(x, \lambda_n, \mu_n) \le 0. \tag{4.24}$$

If $F \in \mathcal{C}_{Lip}(M)$ *then there exists a constant* $\tilde{C}_2$ *such that for any* $n \in \mathbb{N}$,

$$R_n(x, \lambda, \mu) - R(x, \lambda_n, \mu_n) \le \tilde{C}_2 n^{-1/4}(\ln n)^{3/4}. \tag{4.25}$$

Proof. Fix an initial capital $x \ge 0$ and $n \in \mathbb{N}$. There exists a portfolio $\pi = (x, \{\gamma(t)\}_{t=0}^T) \in \mathcal{A}(x, \lambda_n, \mu_n)$ such that

$$R(\pi, \lambda_n, \mu_n) < \frac{1}{n} + R(x, \lambda_n, \mu_n). \tag{4.26}$$

For simplicity we extend the portfolio π to $\mathbb{R}_+$, by setting $\gamma(t) = 0$ for $t > T$, i.e., the portfolio value remains constant after the maturity date T. Set $u_n(k) = \gamma(\theta_k^{(n)})$, $0 \le k \le n$. Define the adapted (to the filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$) process $\{\gamma_n(t)\}_{t=0}^{\infty}$ by

$$\gamma_n(t) = \sum_{k=0}^{n-1} \mathbb{I}_{\theta_k^{(n)}<t\le\theta_{k+1}^{(n)}} u_n(k). \tag{4.27}$$

Consider the portfolio $\tilde{\pi} = (x, \{\gamma_n(t)\}_{t=0}^{\infty})$ in a BS model, for which purchase and sale of the risky asset are subject to proportional transaction costs of rate λ and μ, respectively. Observe that for every $i < n$ we have the inequalities $\exp(2\kappa\sqrt{\frac{T}{n}}) \inf_{\theta_i^{(n)}\le t\le\theta_{i+1}^{(n)}} S(t) \ge S(\theta_{i+1}^{(n)})$ and $\exp(-2\kappa\sqrt{\frac{T}{n}}) \sup_{\theta_i^{(n)}\le t\le\theta_{i+1}^{(n)}} S(t) \le S(\theta_{i+1}^{(n)})$. Thus for any $0 \le i < n$

$$(4.28)\quad (1-\mu_n)\int_{\theta_i^{(n)}}^{\theta_{i+1}^{(n)}} S(t)\,d\gamma^-(t)-(1+\lambda_n)\int_{\theta_i^{(n)}}^{\theta_{i+1}^{(n)}} S(t)\,d\gamma^+(t)$$
$$\le (1-\mu)S(\theta_{i+1}^{(n)})\int_{\theta_i^{(n)}}^{\theta_{i+1}^{(n)}} d\gamma^-(t)-(1+\lambda)S(\theta_{i+1}^{(n)})\int_{\theta_i^{(n)}}^{\theta_{i+1}^{(n)}} d\gamma^+(t)$$
$$\le S(\theta_{i+1}^{(n)})\left((1-\mu)(u_n(i+1)-u_n(i))^- -(1+\lambda)(u_n(i+1)-u_n(i))^+\right)$$
$$\le S\left(\theta_{i+1}^{(n)}\right)\left((1-\mu)\left(u_n^+(i)-u_n^+(i+1)\right)-(1+\lambda)(u_n^-(i)-u_n^-(i+1))\right).$$

Set $u_n(-1)=0$. From (2.8) and (4.28) it follows that for any $k\le n$

$$(4.29)\quad V_{\lambda,\mu}^{\tilde\pi}(\theta_k^{(n)}) = x+(1-\mu)\big(u_n^+(k-1)S(\theta_k^{(n)})$$
$$+\sum_{i=0}^{k-2}(u_n(i+1)-u_n(i))^- S(\theta_{i+1}^{(n)})\big)-(1+\lambda)\big(u_n^-(k-1)S(\theta_k^{(n)})$$
$$+\sum_{i=0}^{k-2}(u_n(i+1)-u_n(i))^+ S(\theta_{i+1}^{(n)})\big)\ge x+(1-\mu_n)\big(u_n^+(k)S(\theta_k^{(n)})$$
$$+\sum_{i=0}^{k-1}\int_{\theta_i^{(n)}}^{\theta_{i+1}^{(n)}} S(t)\,d\gamma^-(t)\big)-(1+\lambda_n)\big(u_n^-(k)S(\theta_k^{(n)})$$
$$+\sum_{i=0}^{k-1}\int_{\theta_i^{(n)}}^{\theta_{i+1}^{(n)}} S(t)\,d\gamma^+(t)\big)= V_{\lambda_n,\mu_n}^{\pi}(\theta_k^{(n)}\wedge T)\ge 0.$$

Thus, $\tilde\pi\in\mathcal{A}^{W,n}(x,\lambda,\mu)$. From (4.21) and (4.23) we obtain that there exists a stopping time $\tau_n\in\mathcal{T}_{0,n}^W$ such that

$$(4.30)\qquad E\big[\big(Y^{W,n}(\tau_n T/n)-V_{\lambda,\mu}^{\tilde\pi}(\theta_{\tau_n}^{(n)})\big)^+\big]\ge R_n(x,\lambda,\mu)-\frac{1}{n}.$$

Clearly $\theta_{\tau_n}^{(n)}\wedge T\in\mathcal{T}_{[0,T]}$, and so $R(\pi,\lambda_n,\mu_n)\ge E[(Y(\theta_{\tau_n}^{(n)}\wedge T)-V_{\lambda_n,\mu_n}^{\pi}(\theta_{\tau_n}^{(n)}\wedge T))^+]$. This together with (4.29) yields

$$(4.31)\qquad R(\pi,\lambda_n,\mu_n)\ge E\big[\big(Y(\theta_{\tau_n}^{(n)}\wedge T)-V_{\lambda,\mu}^{\tilde\pi}(\theta_{\tau_n}^{(n)})\big)^+\big].$$

From (4.26), (4.30), and (4.31) it follows that

$$(4.32)\qquad R_n(x,\lambda,\mu)\le R(x,\lambda_n,\mu_n)+\frac{2}{n}+E\big|Y^{W,n}(\tau_n T/n)-Y(\theta_{\tau_n}^{(n)}\wedge T)\big|.$$

Observe that

$$(4.33)\qquad \big|\tau_n T/n-\theta_{\tau_n}^{(n)}\wedge T\big|\le\max_{0\le k\le n}\left|\theta_k^{(n)}-\frac{kT}{n}\right|,$$

and so from (4.19) the sequence $\{\tau_n T/n-\theta_{\tau_n}^{(n)}\wedge T\}_{n=1}^\infty$ converges to 0 in probability. This together with (4.20) implies that the sequence $\{Y^{W,n}(\tau_n T/n)-Y(\theta_{\tau_n}^{(n)}\wedge T)\}_{n=1}^\infty$ converges to 0 in probability. From (2.4) and the exponential moment estimates in (4.19) it follows that the sequence $\{Y^{W,n}(\tau_n T/n)-Y(\theta_{\tau_n}^{(n)}\wedge T)\}_{n=1}^\infty$ is uniformly integrable. Consequently, it converges to 0 in $L^1(\Omega,P)$, and from (4.32) we get (4.24). Next, let $F\in\mathcal{C}_{Lip}(M)$. From

(4.19), (4.33), and lemma 4.4 in Dolinsky and Kifer (2008) it follows that there exist constants $\hat{C}_8$, $\hat{C}_9$ such that

$$
\begin{aligned}
(4.34)\qquad & E\left|Y^{W,n}(\tau_n T/n) - Y\big(\theta^{(n)}_{\tau_n} \wedge T\big)\right| \le \hat{C}_7 n^{-1/4}(\ln n)^{3/4} \\
&+E\left|Y\big(\theta^{(n)}_{\tau_n}\big) - Y\big(\theta^{(n)}_{\tau_n} \wedge T\big)\right| \le \hat{C}_7 n^{-1/4}(\ln n)^{3/4} + \hat{C}_8 \\
&\times\Big(\big(E\big(\theta^{(n)}_n - T\big)^2\big)^{1/2} + \big(E\big(\theta^{(n)}_n - T\big)^2\big)^{1/4}\Big) \le \hat{C}_9 n^{-1/4}(\ln n)^{3/4}.
\end{aligned}
$$

From (4.32) and (4.34) we obtain (4.25) and the proof is completed. □

From Lemmas 4.2 and 4.3 we obtain

$$
(4.35)\qquad R(x,\lambda,\mu) \ge \limsup_{n\to\infty} R_n(x,\lambda,\mu), \quad \forall x \ge 0,
$$

and for $F \in \mathcal{C}_{Lip}(M)$ there exist constants $\tilde{C}_3$, $\tilde{C}_4$ such that

$$
(4.36)\qquad R_n(x,\lambda,\mu) \le R(x,\lambda,\mu) + \exp\left(\frac{\tilde{C}_3}{(\mu+\lambda)^2}\right) n^{-1/4} + \tilde{C}_4 n^{-1/4}(\ln n)^{3/4}.
$$

Next, let $x \ge 0$ be an initial capital. For any $n \in \mathbb{N}$ let $\pi_n = \pi_n(x,\lambda,\mu) \in \mathcal{A}^{(n)}(x,\lambda,\mu)$ be the optimal portfolio which is given by (3.19)–(3.20). Define $\tilde{\pi}_n = \psi_n(\pi_n) \in \mathcal{A}^{W,n}(x,\lambda,\mu)$, $n \in \mathbb{N}$. Clearly

$$
(4.37)\qquad R(x,\lambda,\mu) \le R\big(\psi^T_n(\pi_n),\lambda,\mu\big), \qquad \forall n \in \mathbb{N},
$$

where recall that $\psi^T_n(\pi_n) \in \mathcal{A}(x,\lambda,\mu)$ is the restriction of the portfolio $\tilde{\pi}_n$ to the interval $[0,T]$. In view of (4.35)–(4.37) we see that in order to complete the proof of Theorems 2.2 and 2.3 it remains to prove the following lemma.

LEMMA 4.4.

$$
(4.38)\qquad \limsup_{n\to\infty} R\big(\psi^T_n(\pi_n),\lambda,\mu\big) - R_n(x,\lambda,\mu) \le 0.
$$

If $F \in \mathcal{C}_{Lip}(M)$ then there exists a constant $\tilde{C}_5$ such that

$$
(4.39)\qquad R\big(\psi^T_n(\pi_n),\lambda,\mu\big) - R_n(x,\lambda,\mu) \le \tilde{C}_5 n^{-1/4}(\ln n)^{3/4}, \qquad \forall n \in \mathbb{N}.
$$

Proof. Fix n. Let $\tau_n \in \mathcal{T}_{[0,T]}$ be such that

$$
(4.40)\qquad R\big(\psi^T_n(\pi_n),\lambda,\mu\big) < \frac{1}{n} + E\big[\big(Y(\tau_n) - V^{\tilde{\pi}_n}_{\lambda,\mu}(\tau_n)\big)^+\big].
$$

Define $\nu_n = n \wedge \min\{k|\theta^{(n)}_k \ge \tau_n\}$. Observe that $\nu_n \in \mathcal{T}^W_{0,n}$. Note that the process $\{(Y^{W,n}(kT/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}(\theta^{(n)}_k))^+\}^n_{k=0}$ is adapted to the filtration $\{\mathcal{G}^n_k\}^n_{k=0}$, thus from standard dynamical programming (see Peskir and Shiryaev 2006) it follows that

$$(4.41)\quad E\big[\big(Y^{W,n}(\nu_n T/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big)^+\big] \le \sup_{\zeta\in\mathcal{T}^W_{0,n}} E\big[\big(Y^{W,n}(\zeta T/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\zeta}\big)\big)^+\big]$$
$$= \sup_{\zeta\in\mathcal{S}_{0,n}} E\big[\big(Y^{W,n}(\zeta T/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\zeta}\big)\big)^+\big].$$

Recall the map Π_n which was introduced after Theorem 2.2. Notice that $\Pi_n : \mathcal{T}_{0,n} \to \mathcal{S}_{0,n}$ is a bijection and for any random variable $U \in L^\infty(\mathcal{F}^\xi_n, P_n)$, $E\Pi_n(U) = E_n U$. From (2.24), (3.15), and (4.18) we obtain

$$(4.42)\quad \sup_{\zeta\in\mathcal{S}_{0,n}} E\big[\big(Y^{W,n}(\zeta T/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\zeta}\big)\big)^+\big] = \sup_{\sigma\in\mathcal{T}_{0,n}} E\big(\Pi_n\big[\big(Y^{(n)}(\sigma) - V^{\pi_n}_{\lambda,\mu}(\sigma)\big)^+\big]\big)$$
$$= \sup_{\sigma\in\mathcal{T}_{0,n}} E_n\big[\big(Y^{(n)}(\sigma) - V^{\pi_n}_{\lambda,\mu}(\sigma)\big)^+\big] = R_n(\pi_n, \lambda, \mu) = R_n(x, \lambda, \mu).$$

From (4.41) and (4.42)

$$(4.43)\quad E\big[\big(Y^{W,n}(\nu_n T/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big)^+\big] \le R_n(x, \lambda, \mu).$$

The portfolio value process $\{V^{\tilde{\pi}_n}_{\lambda,\mu}(t)\}_{t=0}^\infty$ is a supermartingale with respect to the measure $\tilde{P}$. Note that $\theta^{(n)}_{\nu_n} \ge \tau_n \wedge \theta^{(n)}_n$. Thus,

$$(4.44)\quad V^{\tilde{\pi}_n}_{\lambda,\mu}(\tau_n) = V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\tau_n \wedge \theta^{(n)}_n\big) \ge \tilde{E}\big(V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big|\mathcal{F}_{\tau_n\wedge\theta^{(n)}_n}\big).$$

From (2.3), (4.44), and Jensen's inequality it follows that

$$(4.45)\quad E\big[\big(Y\big(\tau_n \wedge \theta^{(n)}_n\big) - V^{\tilde{\pi}_n}_{\lambda,\mu}(\tau_n)\big)^+\big] = \tilde{E}\left(\frac{1}{Z\big(\tau_n \wedge \theta^{(n)}_n\big)}\big(Y\big(\tau_n \wedge \theta^{(n)}_n\big) - V^{\tilde{\pi}_n}_{\lambda,\mu}(\tau_n)\big)^+\right)$$
$$\le \tilde{E}\left(\frac{1}{Z\big(\tau_n \wedge \theta^{(n)}_n\big)}\big(Y\big(\tau_n \wedge \theta^{(n)}_n\big) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big)^+\right)$$
$$= E\left(\frac{Z\big(\theta^{(n)}_{\nu_n}\big)}{Z\big(\tau_n \wedge \theta^{(n)}_n\big)}\big(Y\big(\tau_n \wedge \theta^{(n)}_n\big) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big)^+\right).$$

From (4.40) and (4.45),

$$(4.46)$$
$$R\big(\psi^T_n(\pi_n), \lambda, \mu\big) < \frac{1}{n} + E\big|Y(\tau_n) - Y\big(\tau_n \wedge \theta^{(n)}_n\big)\big| + E\left(\left|\frac{Z\big(\theta^{(n)}_{\nu_n}\big)}{Z\big(\tau_n \wedge \theta^{(n)}_n\big)} - 1\right| \sup_{0\le t\le T} Y(t)\right)$$
$$+ E\big[\big(Y\big(\tau_n \wedge \theta^{(n)}_n\big) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big)^+\big]$$
$$\le \frac{1}{n} + E\big[\big(Y^{W,n}(\nu_n T/n) - V^{\tilde{\pi}_n}_{\lambda,\mu}\big(\theta^{(n)}_{\nu_n}\big)\big)^+\big] + \Upsilon^{(1)}_n + \Upsilon^{(2)}_n + \Upsilon^{(3)}_n,$$

where

$$\Upsilon^{(1)}_n = E\big(\big|Z\big(\theta^{(n)}_{\nu_n}\big) - Z\big(\tau_n \wedge \theta^{(n)}_n\big)\big| \sup_{0\le t\le T} Z^{-1}(t) \sup_{0\le t\le T} Y(t)\big),$$
$$\Upsilon^{(2)}_n = E\big(\big|Y(\tau_n) - Y\big(\tau_n \wedge \theta^{(n)}_n\big)\big| + \big|Y\big(\tau_n \wedge \theta^{(n)}_n\big) - Y\big(\theta^{(n)}_{\nu_n}\big)\big|\big)$$
$$\text{and}\quad \Upsilon^{(3)}_n = E\big|Y\big(\theta^{(n)}_{\nu_n}\big) - Y^{W,n}(\nu_n T/n)\big|.$$

It follows from the definitions that $\tau_n - \tau_n \wedge \theta_n^{(n)} \leq |T - \theta_n^{(n)}|$ and $\theta_{\nu_n}^{(n)} - \tau_n \wedge \theta_n^{(n)} \leq \max_{0 \leq k < n} \theta_{k+1}^{(n)} - \theta_k^{(n)} \leq \frac{T}{n} + 2 \max_{1 \leq k \leq n} |\theta_k^{(n)} - \frac{kT}{n}|$. From (4.19) we get that there exists a constant $\hat{C}_{10}$ such that

$$E\big(\max\big(\tau_n - \tau_n \wedge \theta_n^{(n)}, \theta_{\nu_n}^{(n)} - \tau_n \wedge \theta_n^{(n)}\big)^2\big) \leq \frac{\hat{C}_{10}}{n}. \tag{4.47}$$

From (4.19), (4.47), Itô's formula, Itô's isometry, and the Cauchy–Schwarz inequality we get

(4.48)

$$\begin{aligned} E\big(\big(Z(\theta_{\nu_n}^{(n)}) - Z(\tau_n \wedge \theta_n^{(n)})\big)^2\big) &= \left(\frac{\vartheta}{\kappa}\right)^2 E \int_{\tau_n \wedge \theta_n^{(n)}}^{\theta_{\nu_n}^{(n)}} Z^2(t)\, dt \\ &\leq \left(\frac{\vartheta}{\kappa}\right)^2 E\Big(\big(\theta_{\nu_n}^{(n)} - \tau_n \wedge \theta_n^{(n)}\big) \sup_{0 \leq t \leq \theta_n^{(n)} \vee T} Z^2(t)\Big) \leq \hat{C}_{11} n^{-1/2} \end{aligned}$$

for some constant $\hat{C}_{11}$. From (4.48) and the Cauchy–Schwarz inequality we conclude that there exists a constant $\hat{C}_{12}$ such that

$$\Upsilon_n^{(1)} \leq \hat{C}_{12} n^{-1/4}. \tag{4.49}$$

By using similar arguments to those which appear after (4.33) we obtain that $\lim_{n\to\infty} \Upsilon_n^{(2)} = \lim_{n\to\infty} \Upsilon_n^{(3)} = 0$. Thus from (4.43), (4.46), and (4.49) we get (4.38). Finally, let $F \in \mathcal{C}_{Lip}(M)$. From (4.19), (4.47), and lemma 4.4 in Dolinsky and Kifer (2008) we obtain that there exists a constant $\hat{C}_{13}$ such that $\Upsilon_n^{(2)} + \Upsilon_n^{(3)} \leq \hat{C}_{13} n^{-1/4} (\ln n)^{3/4}$. This together with (4.43), (4.46), and (4.49) gives (4.39). $\square$

5. PROOF OF THEOREM 2.1

In this section we assume that the parameters x, λ, μ are fixed. Let $\mathcal{I} \subset [0, T]$ be a dense set in $[0, T]$ and let $\mathcal{T}_\mathcal{I} \subset \mathcal{T}_{[0,T]}$ be the set of all stopping times with a finite number of values which belong to $\mathcal{I}$.

LEMMA 5.1. *For any* $\pi \in \mathcal{A}(x, \lambda, \mu)$,

$$R(\pi, \lambda, \mu) = \sup_{\tau \in \mathcal{T}_\mathcal{I}} E\big[\big(Y(\tau) - V_{\lambda,\mu}^\pi(\tau)\big)^+\big]. \tag{5.1}$$

Proof. Clearly $R(\pi, \lambda, \mu) \geq \sup_{\tau \in \mathcal{T}_\mathcal{I}} E[(Y(\tau) - V_{\lambda,\mu}^\pi(\tau))^+]$. Thus it is sufficient to show that $R(\pi, \lambda, \mu) \leq \sup_{\tau \in \mathcal{T}_\mathcal{I}} E[(Y(\tau) - V_{\lambda,\mu}^\pi(\tau))^+]$. Choose $\epsilon > 0$. There exists $\tilde{\tau} \in \mathcal{T}_{[0,T]}$ such that

$$R(\pi, \lambda, \mu) < E\big[\big(Y(\tilde{\tau}) - V_{\lambda,\mu}^\pi(\tilde{\tau})\big)^+\big] + \epsilon. \tag{5.2}$$

For any $n \in \mathbb{N}$ there exists a finite set $\mathcal{I}_n \subset \mathcal{I}$ for which $\bigcup_{z \in \mathcal{I}_n} (z - \frac{1}{n}, z + \frac{1}{n}) \supseteq [0, T]$. Let a_n be the maximal element of $\mathcal{I}_n$. Define $\tau_n = \min\{t \in \mathcal{I}_n | t \geq \tilde{\tau}\} \mathbb{I}_{\tilde{\tau} \leq a_n} + a_n \mathbb{I}_{\tilde{\tau} > a_n}$. Clearly, $\tau_n \leq a_n$ a.s. and for $t \in \mathcal{I}_n \setminus \{a_n\}$ we have $\{\tau_n \leq t\} = \{\tilde{\tau} \leq t\} \in \mathcal{F}_t$. Thus $\tau_n \in \mathcal{T}_\mathcal{I}$. Furthermore, $|\tau_n - \tilde{\tau}| \leq \frac{2}{n}$ and so $\tau_n \to \tilde{\tau}$ a.s. From (2.8) it follows that the stochastic process $\{V_{\lambda,\mu}^\pi(t)\}_{t=0}^T$ is left continuous with right-hand limits and has only negative jumps (in the discontinuity points). Thus $V^{\pi}_{\lambda,\mu}(\tilde{\tau}) \geq \limsup_{n\to\infty} V^{\pi}_{\lambda,\mu}(\tau_n)$ a.s. By using (5.2) and Fatou's lemma we obtain that

$$(5.3)\ R(\pi,\lambda,\mu) < \epsilon + E\big[\liminf_{n\to\infty}\big(Y(\tau_n) - V^{\pi}_{\lambda,\mu}(\tau_n)\big)^+\big]$$
$$\leq \epsilon + \liminf_{n\to\infty} E\big[\big(Y(\tau_n) - V^{\pi}_{\lambda,\mu}(\tau_n)\big)^+\big] \leq \epsilon + \sup_{\tau\in\mathcal{T}_{\mathcal{I}}} E\big[\big(Y(\tau) - V^{\pi}_{\lambda,\mu}(\tau)\big)^+\big]$$

and the result follows by letting $\epsilon\downarrow 0$. □

Next, let $\{\hat{\pi}_n = (x, \hat{\gamma}_n)\}_{n=1}^{\infty} \subset \mathcal{A}(x,\lambda,\mu)$ be a sequence such that

$$(5.4)\qquad \lim_{n\to\infty} R(\hat{\pi}_n,\lambda,\mu) = R(x,\lambda,\mu).$$

From (4.3) and Lemma 4.1 we obtain that $conv\big\{\int_0^T S(u)|d\hat{\gamma}_n|(u)\big\}_{n=1}^{\infty}$ is bounded in $L^0(P)$. This together with lemma 3.1 in Guasoni (2002b) yields that the set $conv\big\{\int_0^T |d\hat{\gamma}_n|(u)\big\}_{n=1}^{\infty}$ is also bounded in $L^0(P)$. From lemma 3.4 in Guasoni (2002b) there exists a sequence $\eta_n \in conv(\hat{\gamma}_n, \hat{\gamma}_{n+1}, \ldots)$ such that η_n converges a.s. in $dtdP$ to a finite variation process. In fact, from the proof of this lemma we get a stronger result. We obtain that there exist nondecreasing, left-continuous adapted processes $\{\alpha(t)\}_{t=0}^T$ and $\{\beta(t)\}_{t=0}^T$ with $\alpha(0) = \beta(0) = 0$, such that

$$(5.5)\qquad \lim_{n\to\infty}\eta_n^+ = \alpha \quad\text{and}\quad \lim_{n\to\infty}\eta_n^- = \beta, \quad \text{a.s in } dtdP,$$

where

$$(5.6)\qquad \eta_n^+(t) = \frac{\eta_n(t) + \int_0^t |d\eta_n|(s)}{2} \quad\text{and}$$
$$\eta_n^-(t) = \frac{\int_0^t |d\eta_n|(s) - \eta_n(t)}{2}, \quad t\in[0,T], \quad n\in\mathbb{N}.$$

In particular, there exists a countable dense set $\mathcal{I} \subset [0,T]$, such that $0\in\mathcal{I}$ and

$$(5.7)\ P\Big\{\lim_{n\to\infty}\eta_n^+(t) = \alpha(t), \forall t\in\mathcal{I}\Big\} = 1 \quad\text{and}\quad P\Big\{\lim_{n\to\infty}\eta_n^-(t) = \beta(t), \forall t\in\mathcal{I}\Big\} = 1.$$

Define $\gamma = \alpha - \beta$. Clearly, γ is an adapted process of bounded variation with left-continuous paths and $\gamma(0) = 0$. Finally, we prove that $\pi \coloneqq (x,\gamma)$ is an optimal portfolio, i.e., $\pi\in\mathcal{A}(x,\lambda,\mu)$ and $R(\pi,\lambda,\mu) = R(x,\lambda,\mu)$. From (2.8) and the integration by parts formula we obtain that for any portfolio $\tilde{\pi} = (x,\tilde{\gamma}) \in \mathcal{A}(x,\lambda,\mu)$

$$(5.8)\qquad V^{\tilde{\pi}}_{\lambda,\mu}(t) = x + \int_0^t \tilde{\gamma}(t)\,dS(t) - \mu\left(\int_0^t S(u)\,d\tilde{\gamma}^-(u) + \tilde{\gamma}(t)^+S(t)\right)$$
$$-\lambda\left(\int_0^t S(u)\,d\tilde{\gamma}^+(u) + \tilde{\gamma}(t)^-S(t)\right), \quad t\in[0,T].$$

For any $n\in\mathbb{N}$ there exists $m\in\mathbb{N}$, natural numbers $N_1,\ldots,N_m \geq n$, positive numbers $\lambda_1,\ldots,\lambda_m > 0$ such that $\sum_{i=1}^m \lambda_i = 1$ and $\eta_n = \sum_{i=1}^m \lambda_i\hat{\gamma}_{N_i}$. From (5.8) it follows that for the portfolio $\bar{\pi}_n := (x,\eta_n)$, we have

$$V^{\bar{\pi}_n}_{\lambda,\mu}(t) \geq \sum_{i=1}^{m} \lambda_i V^{\hat{\pi}_{N_i}}_{\lambda,\mu}(t), \quad t \in [0, T].$$

Thus $\bar{\pi}_n \in \mathcal{A}(x, \lambda, \mu)$. The shortfall risk measure $R(\cdot, \lambda, \mu)$ is a convex functional of the wealth process $V^{\cdot}_{\lambda,\mu}$, and so we conclude

$$R(\bar{\pi}_n, \lambda, \mu) \leq \sup_{k \geq n} R(\hat{\pi}_k, \lambda, \mu). \tag{5.9}$$

From (5.4) and (5.9),

$$\lim_{n\to\infty} R(\bar{\pi}_n, \lambda, \mu) = R(x, \lambda, \mu). \tag{5.10}$$

From (5.7), and theorem 12.16 in Protter and Morrey (1991),

$$\begin{aligned} \int_0^t S(u)\, d\alpha(u) &= \lim_{n\to\infty} \int_0^t S(u)\, d\eta_n^+(u) \quad \text{and} \\ \int_0^t S(u)\, d\beta(u) &= \lim_{n\to\infty} \int_0^t S(u)\, d\eta_n^-(u), \quad \text{a.s. } \forall t \in \mathcal{I}. \end{aligned} \tag{5.11}$$

Thus,

$$\begin{aligned} &\int_0^t S(u)\, d\gamma(u) = \lim_{n\to\infty} \int_0^t S(u)\, d\eta_n(u) \ \text{ and } \int_0^t S(u)|d\gamma|(u) \\ &\qquad \leq \int_0^t S(u)\, d\alpha(u) + \int_0^t S(u)\, d\beta(u) = \lim_{n\to\infty} \int_0^t S(u)|d\eta_n|(u), \quad \text{a.s. } \forall t \in \mathcal{I}. \end{aligned} \tag{5.12}$$

This together with (2.7) and (2.8) gives

$$V^{\pi}_{\lambda,\mu}(t) \geq \lim_{n\to\infty} V^{\bar{\pi}_n}_{\lambda,\mu}(t) \geq 0, \quad \forall t \in \mathcal{I}, \tag{5.13}$$

and so $\pi \in \mathcal{A}(x, \lambda, \mu)$. By combining Fatou's lemma together with Lemma 5.1, (5.10), and (5.13) we obtain

$$\begin{aligned} R(\pi, \lambda, \mu) &= \sup_{\tau \in \mathcal{T}_{\mathcal{I}}} E\Big[\big(Y(\tau) - V^{\pi}_{\lambda,\mu}(\tau)\big)^+\Big] \leq \sup_{\tau \in \mathcal{T}_{\mathcal{I}}} E\big[\lim_{n\to\infty} \big(Y(\tau) - V^{\bar{\pi}_n}_{\lambda,\mu}(\tau)\big)^+\big] \\ &\leq \sup_{\tau \in \mathcal{T}_{\mathcal{I}}} \lim\inf_{n\to\infty} E\big[\big(Y(\tau) - V^{\bar{\pi}_n}_{\lambda,\mu}(\tau)\big)^+\big] \leq \lim_{n\to\infty} R(\bar{\pi}_n, \lambda, \mu) = R(x, \lambda, \mu). \end{aligned}$$

Thus, $R(\pi, \lambda, \mu) = R(x, \lambda, \mu)$ and the proof is completed. □

6. STABILITY RESULT FOR SMALL TRANSACTION COSTS

In this section we show that in the BS model, when the transaction costs tend to 0 the corresponding shortfall risks converge to the shortfall risk in the complete BS market.

Consider the BS model in the absence of transaction costs (complete market). In this case a self-financing strategy π with an initial capital x is a pair $(x, \{\gamma(t)\}_{t=0}^T)$, such that the process $\{\gamma(t)\}_{t=0}^T$ is progressively measurable with respect to the filtration $\mathcal{F}_t$, $t \geq 0$ and satisfies

$$\int_0^T \big(\gamma(t)S(t)\big)^2 dt < \infty \quad \text{a.s.} \tag{6.1}$$

The wealth process $\{V^\pi(t)\}_{t=0}^T$ for a strategy $\pi = (x, \{\gamma(t)\}_{t=0}^T)$ is given by

$$V^\pi(t) = x + \int_0^t \gamma(u)\, dS(u), \qquad \forall t \in [0, T]. \tag{6.2}$$

A self-financing strategy π is called *admissible* if $V^\pi(t) \geq 0$ for all $t \in [0, T]$ and the set of such strategies with an initial capital x will be denoted by $\mathcal{A}(x)$. The shortfall risk is defined by

$$\begin{aligned} R(\pi) &= \sup_{\tau \in \mathcal{T}_{[0,T]}} E[(Y(\tau) - V^\pi(\tau))^+] \text{ and} \\ R(x) &= \inf_{\pi \in \mathcal{A}(x)} \sup_{\tau \in \mathcal{T}_{[0,T]}} E[(Y(\tau) - V^\pi(\tau))^+]. \end{aligned} \tag{6.3}$$

THEOREM 6.1. *For any initial capital $x \geq 0$,*

$$R(x) = \lim_{\lambda \downarrow 0} \lim_{\mu \downarrow 0} R(x, \lambda, \mu). \tag{6.4}$$

Proof. For $x = 0$ the statement is trivial. Fix an initial capital $x > 0$. Observe that the function $R(x, \lambda, \mu)$ is a nondecreasing function with respect to the parameters λ and μ, hence the limit in the right-hand side of (6.4) exists. Clearly,

$$R(x) \leq \lim_{\lambda \downarrow 0} \lim_{\mu \downarrow 0} R(x, \lambda, \mu). \tag{6.5}$$

Next we show that $R(x) \geq \lim_{\lambda \downarrow 0} \lim_{\mu \downarrow 0} R(x, \lambda, \mu)$. Choose $\epsilon > 0$. For any $z \geq 0$ let $\mathcal{A}^K(z) \subset \mathcal{A}(z)$ be the subset consisting of all $\pi \in \mathcal{A}(z)$, such that the portfolio process $\{V^\pi(t)\}_{t=0}^T$ is a right continuous martingale with respect to the martingale measure $\tilde{P}$, and $V^\pi(T) = f(W(t_1), \dots, W(t_k))$ for some $k \in \mathbb{N}$, $t_1, \dots, t_k \in [0, T]$ and a smooth function $f \in C_0^\infty(\mathbb{R}^k)$ with a compact support. Using the same arguments as in lemmas 4.1–4.3 in Dolinsky and Kifer (2008) we obtain that there exists $z < x$ and $\tilde{\pi} \in \mathcal{A}^K(z)$ such that

$$R(\tilde{\pi}) < R(x) + \epsilon. \tag{6.6}$$

From the Itô formula it follows that there exists a *càdlàg*-adapted process $\{\eta(t)\}_{t=0}^T$ such that

$$V^{\tilde{\pi}}(t) = z + \int_0^t \eta(u)\, dS(u), \qquad \forall t \in [0, T]. \tag{6.7}$$

From lemma A.3 in Levental and Skorohod (1997) we obtain that there exists a "simple process" $\{\tilde{\eta}(t)\}_{t=0}^T$ such that

$$\sup_{0 \leq t \leq T} \left| \int_0^t \eta(u)\, dS(u) - \int_0^t \tilde{\eta}(u)\, dS(u) \right| < x - z \quad \text{a.s.} \tag{6.8}$$

A "simple process" means (see definition A.4 in Levental and Skorohod 1997) that

$$\tilde{\eta}(t) = \sum_{n=0}^\infty u_n \mathbb{I}_{\sigma_n \leq t < \sigma_{n+1}}, \qquad t \in [0, T], \tag{6.9}$$

where $0 = \sigma_0 \leq \sigma_1 \leq \dots \sigma_n \leq \dots$ is a sequence of stopping times with values in the set $[0, T]$, such that the set $\{n \in \mathbb{N} | \sigma_n = T\}$ is not empty a.s., and for any n, u_n is $\mathcal{F}_{\sigma_n}$ measurable. Define the process $\{\gamma(t)\}_{t=0}^T$ by

$$\gamma(t) = \sum_{n=0}^{\infty} u_n \mathbb{I}_{\sigma_n < t \leq \sigma_{n+1}}, \qquad t \in [0, T]. \tag{6.10}$$

Clearly $\gamma(0) = 0$ and γ is an adapted process of bounded variation with left-continuous paths. Consider the portfolio $\pi = (x, \{\gamma(t)\}_{t=0}^T)$. From (6.7)–(6.10) we obtain that for any stopping time $\tau \in \mathcal{T}_{[0,T]}$

$$\begin{aligned} V^\pi(\tau) &= x + \int_0^\tau \gamma(u)\, dS(u) = x + \int_0^\tau \tilde{\eta}(u)\, dS(u) \\ &\geq z + \int_0^\tau \eta(u)\, dS(u) = V^{\tilde{\pi}}(\tau) \geq 0. \end{aligned} \tag{6.11}$$

From (6.6) and (6.11)

$$R(\pi) < R(x) + \epsilon. \tag{6.12}$$

Set

$$\delta_q^{(n)} = \frac{(1-q)x}{n}, \qquad n \in \mathbb{N},\ 0 < q < 1. \tag{6.13}$$

We assume that $n > x$ and so $\delta_q^{(n)} < 1$. Introduce the stopping times

$$\tau_n = T \wedge \inf\left\{t \middle| \int_0^t S(u)|d\gamma|(u) + |\gamma(t)|S(t) \geq n\right\}, \quad n \in \mathbb{N}. \tag{6.14}$$

The stochastic process $\left\{\int_0^t S(u)|d\gamma|(u) + |\gamma(t)|S(t)\right\}_{t=0}^T$ is left continuous, and so for any $t \leq T$,

$$\int_0^{t\wedge\tau_n} S(u)|d\gamma|(u) + |\gamma(t\wedge\tau_n)|S(t\wedge\tau_n) \leq n. \tag{6.15}$$

Notice that

$$\lim_{n\to\infty} \tau_n = T \quad \text{a.s.} \tag{6.16}$$

For any $n \in \mathbb{N}$ and $0 < q < 1$, $\{q\gamma(t)\mathbb{I}_{t\leq\tau_n}\}_{t=0}^T$ is an adapted process of bounded variation with left-continuous paths. Consider the portfolio $\pi_q^{(n)} = (x, \{q\gamma(t)\mathbb{I}_{t\leq\tau_n}\}_{t=0}^T)$. From (2.8), (6.15), and the integration by parts formula we obtain that

$$\begin{aligned} V^{\pi_q^{(n)}}_{\delta_q^{(n)},\delta_q^{(n)}}(t) &= V^{\pi_q^{(n)}}_{\delta_q^{(n)},\delta_q^{(n)}}(t\wedge\tau_n) = qx + (1-q)x + q\int_0^{t\wedge\tau_n} \gamma(u)\, dS(u) \\ &\quad - q\delta_q^{(n)}\left(\int_0^{t\wedge\tau_n} S(u)|d\gamma|(u) + |\gamma(t\wedge\tau_n)|S(t\wedge\tau_n)\right) \\ &\geq qx + q\int_0^{t\wedge\tau_n} \gamma(u)\, dS(u) = qV^\pi(t\wedge\tau_n) \geq 0. \end{aligned} \tag{6.17}$$

We conclude that $\pi_q^{(n)} \in \mathcal{A}(x, \delta_q^{(n)}, \delta_q^{(n)})$. From (6.12) and (6.17) we get

(6.18)
$$\begin{aligned}
R\big(x, \delta_q^{(n)}, \delta_q^{(n)}\big) &\le R\big(\pi_q^{(n)}, \delta_q^{(n)}, \delta_q^{(n)}\big) \le \sup_{\tau\in\mathcal{T}_{[0,T]}} E[(Y(\tau) - qV^{\pi}(\tau\wedge\tau_n))^+] \\
&\le q \sup_{\tau\in\mathcal{T}_{[0,T]}} E[(Y(\tau) - V^{\pi}(\tau\wedge\tau_n))^+] + (1-q)E\sup_{0\le t\le T} Y(t) \le \epsilon + R(x) \\
&\quad + E\sup_{0\le t\le T}|Y(t) - Y(t\wedge\tau_n)| + (1-q)E\sup_{0\le t\le T} Y(t).
\end{aligned}$$

Since $\lim_{\lambda\downarrow 0}\lim_{\mu\downarrow 0} R(x, \lambda, \mu)$ exists (see the argument before (6.5)), then from the equality $\lim_{q\uparrow 1}\delta_q^{(n)} = 0$ we obtain that for any $n \in \mathbb{N}$,

$$\lim_{\lambda\downarrow 0}\lim_{\mu\downarrow 0} R(x, \lambda, \mu) = \lim_{q\uparrow 1} R\big(x, \delta_q^{(n)}, \delta_q^{(n)}\big).$$

Therefore, (6.18) implies that for any n,

$$\lim_{\lambda\downarrow 0}\lim_{\mu\downarrow 0} R(x, \lambda, \mu) \le \epsilon + R(x) + E\sup_{0\le t\le T}|Y(t) - Y(t\wedge\tau_n)|.$$

We conclude that

$$\lim_{\lambda\downarrow 0}\lim_{\mu\downarrow 0} R(x, \lambda, \mu) \le \epsilon + R(x) + \lim\inf_{n\to\infty} E\sup_{0\le t\le T}|Y(t) - Y(t\wedge\tau_n)|. \tag{6.19}$$

From (6.16) we obtain that $\lim_{n\to\infty} E\sup_{0\le t\le T}|Y(t) - Y(t\wedge\tau_n)| = 0$. This together with (6.19) completes the proof. □

Let us notice that by similar arguments to the above one can prove that the function $R(x, \lambda, \mu)$ is continuous in λ and μ.

REMARK 6.2. Consider an American call option $Y(t) = (S(t) - Ke^{-rt})^+$, $t \le T$ with parameters $K, r > 0$. We define a buy-and-hold strategy with an initial capital x as a strategy of the form $\gamma \equiv \frac{x}{s_0(1+\lambda)}$, namely we use all the initial capital in order to buy as much stocks as possible at time $t = 0$. For such a portfolio, the shortfall risk is given by $\sup_{\tau\in\mathcal{T}_{[0,T]}} E[(Y(\tau) - \frac{(1-\mu)S(\tau)}{s_0(1+\lambda)})^+]$, which is bigger than 0 if $x < \frac{s_0(1+\lambda)}{1-\mu}$. Clearly $V^* = \tilde{E}Y(T) < s_0$ is the price of the above call option in the complete BS model. From Theorem 6.1 it follows that $\lim_{\lambda\downarrow 0}\lim_{\mu\downarrow 0} R(V^*, \lambda, \mu) = 0$. In particular we obtain that in the presence of transaction costs, for an initial capital $x \in [V^*, s_0)$ and for sufficiently small $\lambda, \mu > 0$ the buy-and-hold strategies are not optimal (unlike for the superhedging case) for the shortfall risk measure. In Section 7 we provide some numerical examples in the binomial models, and we compare the buy-and-hold strategy with the optimal one.

REMARK 6.3. The above-mentioned limit theorems can be extended to the setting of game options, which were introduced in Kifer (2000). Let $F_1, F_2 \in \mathcal{C}(M)$ such that $F_1 \ge F_2$. Consider the game options with the following payoffs.

$$\begin{aligned}
&H(t, s) = F_1(t, S)\mathbb{I}_{t<s} + F_2(s, S)\mathbb{I}_{s\le t}, \quad t, s \in [0, T] \text{ and} \\
&H^{(n)}(k, l) = F_1\left(\frac{kT}{n}, S^{(n)}\right)\mathbb{I}_{k<l} + F_2\left(\frac{lT}{n}, S^{(n)}\right)\mathbb{I}_{l\le k}, \quad n \in \mathbb{N},\ 0 \le k, l \le n.
\end{aligned} \tag{6.20}$$

The terms $H(t, s)$ and $H^{(n)}(k, l)$, are the payoff functions for the BS model and the n-step binomial model, respectively. For game options in the BS model with transaction costs λ, μ a hedge with an initial capital x is a pair $(\pi, \sigma) \in \mathcal{A}(x, \lambda, \mu) \times \mathcal{T}_{[0,T]}$. The shortfall risk in the BS model is defined by

$$R^{(g)}(\pi, \sigma, \lambda, \mu) = \sup_{\tau \in \mathcal{T}_{[0,T]}} E\big[\big(H(\sigma, \tau) - V^{\pi}_{\lambda,\mu}(\sigma \wedge \tau)\big)^{+}\big] \text{ and}$$

$$R^{(g)}(x, \lambda, \mu) = \inf_{\pi \in \mathcal{A}(x,\lambda,\mu)} \inf_{\sigma \in \mathcal{T}_{[0,T]}} \sup_{\tau \in \mathcal{T}_{[0,T]}} E\big[\big(H(\sigma, \tau) - V^{\pi}_{\lambda,\mu}(\sigma \wedge \tau)\big)^{+}\big].$$

For the binomial models we have analogous definitions. It can be shown that Theorems 2.2, 2.3, and 6.1 can be extended to this case. The extension of Theorem 2.1 is more complicated and requires new tools. In the proof of Theorem 2.1 we used heavily the convexity of the shortfall risk as a functional of the wealth process. For game options the functional

$$R^{(g)}(\pi, \lambda, \mu) := \inf_{\sigma \in \mathcal{T}_{[0,T]}} \sup_{\tau \in \mathcal{T}_{[0,T]}} E\big[\big(H(\sigma, \tau) - V^{\pi}_{\lambda,\mu}(\sigma \wedge \tau)\big)^{+}\big]$$

is not convex in the wealth process $V^{\pi}_{\lambda,\mu}$. For now, the question whether there exists an optimal hedge for game options in the BS model remains open. Recently (see Dolinsky 2012), we proved that when dealing with super-replication of game options in the presence of transaction costs, the super-replication price is the cheapest cost of a trivial perfect hedge. For game options a trivial hedge is a pair which consists of a buy-and-hold strategy and a hitting time of the stock process into a Borel set.

7. NUMERICAL RESULTS

In this section we provide some numerical results for the shortfall risk in the binomial models. The parameters of the BS model are $s_0 = 1$, $\sigma = 0.25$, $\vartheta = 0.125$, and $T = 1$. We consider a call options with the strike price $K = 0.75$. We also assume that the interest rate of our market is $r = 0.02$. Thus the call option price process in the BS model is given by

$$Y(t) = (S(t) - K \exp(-rt))^{+}, \quad t \in [0, 1], \tag{7.1}$$

and for the n-step binomial model by

$$Y^{(n)}(k) = (S^{(n)}(k/n) - K \exp(-rk/n))^{+}, \quad 0 \leq k \leq n. \tag{7.2}$$

For the Markov case the dynamical programming algorithm, which is given by (3.16) and (3.17), can be rewritten in the following more simple way. Define a sequence of functions $\mathcal{J}^{(n)}_k : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$, $k = 0, 1, \ldots, n$ by backward relations

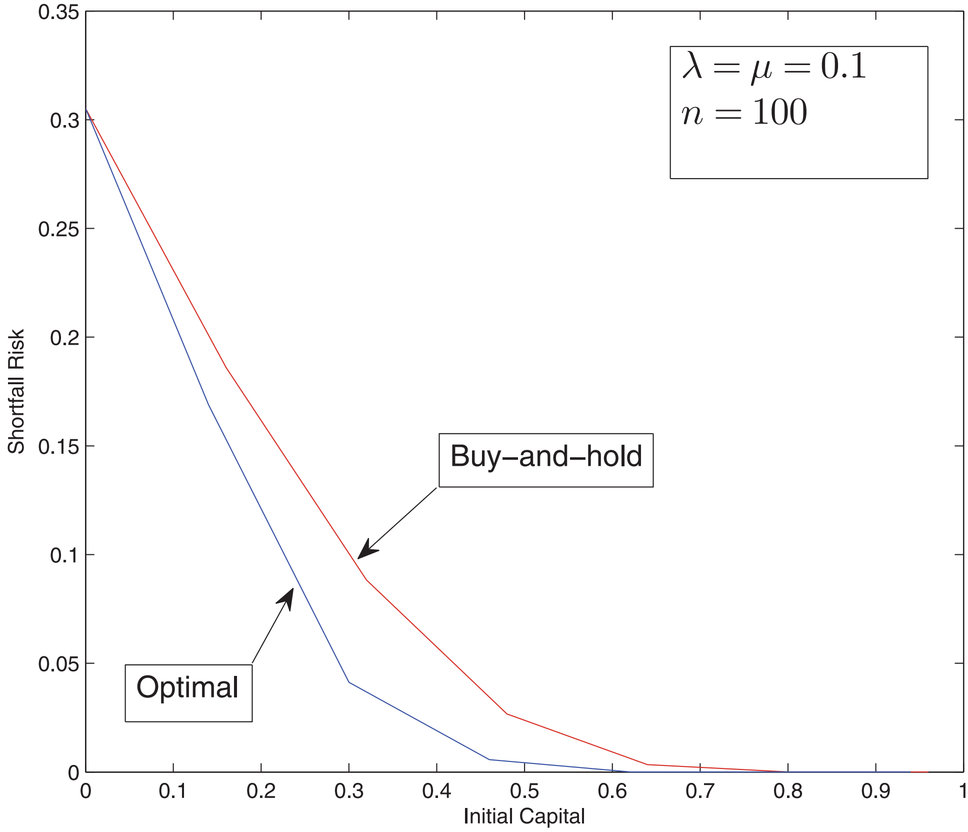


FIGURE 7.1. Shortfall risk for a fixed binomial model 1.

$$\mathcal{J}_n^{(n)}(u, v, z) = \left((z - K\exp(-r))^+ - u\right)^+, \tag{7.3}$$

and for $k < n$

(7.4)

$$\mathcal{J}_k^{(n)}(u, v, z) = \max\Bigg(\left((z - K\exp(-rk/n))^+ - u\right)^+, \inf_{w\in\mathcal{A}_{a_n,b_n}(u,v)} \Big[p^{(n)}\mathcal{J}_{k+1}^{(n)}\left(G(u, v, w, 1+b_n), (1+b_n)(v+w), (1+b_n)z\right) + (1-p^{(n)})\mathcal{J}_{k+1}^{(n)}(G(u, v, w, 1-a_n), (1-a_n)(v+w), (1-a_n)z)\Big]\Bigg),$$

where recall that $p^{(n)}$ was defined after (2.14) and $a_n, b_n, G, \mathcal{A}(a, b)$ were defined in Section 3. Observe that the term $\mathcal{J}_k^{(n)}(u, v, z)$ represents the shortfall risk that the investor is facing (in the n-step binomial model) at time k where u is his portfolio value, v is the wealth that is invested in stocks, and z is the stock price at the moment. Thus,

$$R_n(x, \lambda, \mu) = \mathcal{J}_0^{(n)}(x, 0, s_0), \quad n \in \mathbb{N}. \tag{7.5}$$

Limit Theorems under Transaction Costs

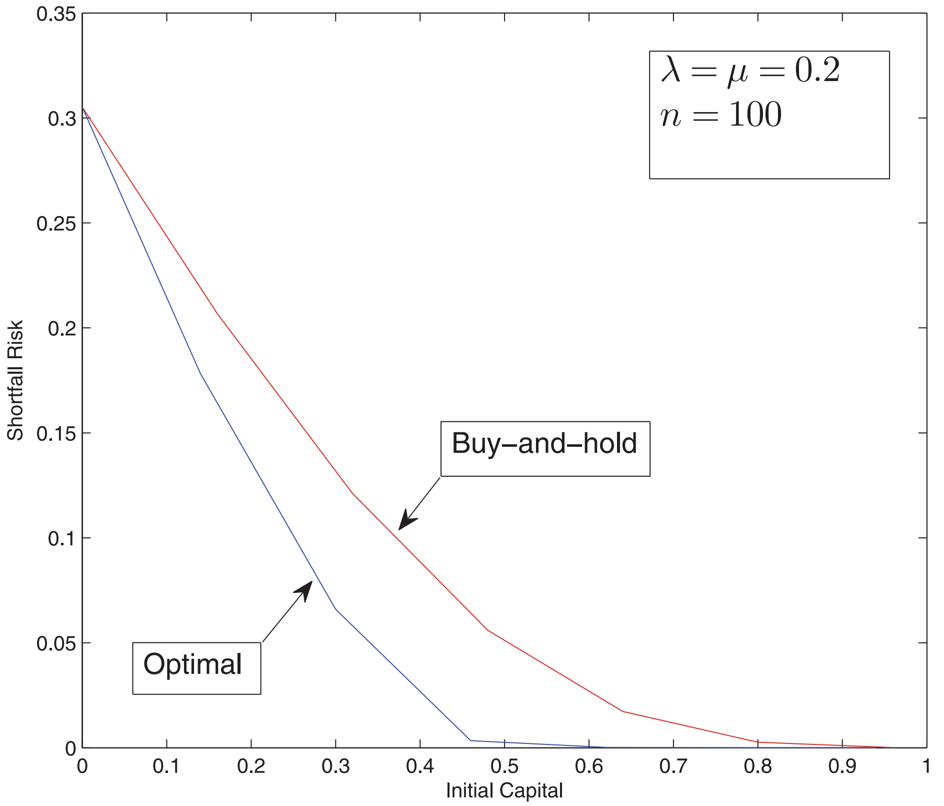


FIGURE 7.2. Shortfall risk for a fixed binomial model 2.

By considering a three-dimensional grid, we constructed a discretization procedure for the dynamical programming given by (7.3) and (7.4). First we compared the optimal shortfall risk with the shortfall risk, which is achieved by a buy-and-hold strategy as defined in Remark 6.2. We considered the 100-step binomial model for different values of the transaction costs parameters λ, μ. In Figure 7.1, for $\lambda = \mu = 0.1$, we plot (on the same graph) the optimal shortfall risk as a function of the initial capital x together with the shortfall risk, which is achieved for the same initial capital if we use the buy-and-hold strategy. In Figure 7.2 we plot the same functions for $\lambda = \mu = 0.2$. Our conclusions from the numerical results are that for partial hedging the buy-and-hold strategy is far from being optimal. In order to minimize the shortfall risk the investor should use dynamic hedging.

The second question that we study is the convergence rate of the functions $R_n(x, \lambda, \mu)$. In Figure 7.3 for $\lambda = \mu = 0.15$ we plot the functions $R_{40}(x, \lambda, \mu)$, $R_{80}(x, \lambda, \mu)$, and $R_{120}(x, \lambda, \mu)$. It seems that the numerical results indicate that the convergence rate is faster than the one we provided in Theorem 2.2. These numerical results give us motivation to find a new tool, which will provide better estimates than those obtained in Theorem 2.2. We emphasize that the Skorohod embedding does not yield error estimates of a rate better than $n^{-1/4}$.

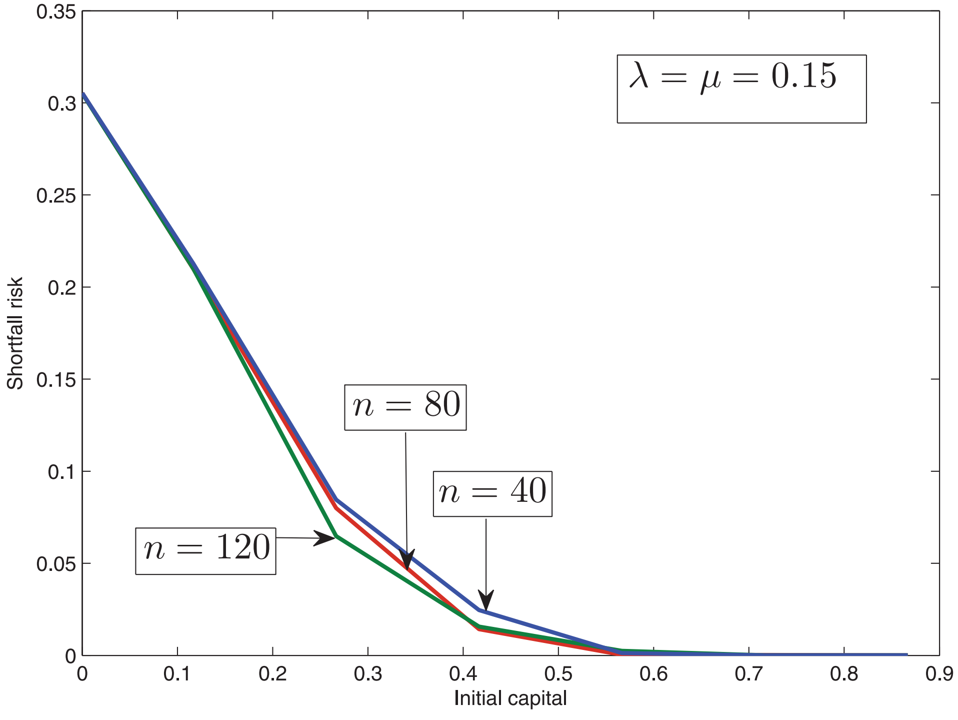


FIGURE 7.3. Convergence rate of binomial models.

REFERENCES

CVITANIC, J., and I. KARATZAS (1996): Hedging and Portfolio Optimization under Transaction Costs, *Math. Finance* 6, 133–165.

CVITANIC, J., H. PHAM, and N. TOUZI (1999): A Closed-Form Solution to the Problem of Superreplication under Transaction Costs, *Finance Stoch.* 3, 35–54.

DOLINSKY, Y. (2012): Hedging of Game Options with the Presence of Transaction Costs, arXiv:1103.1165, v3, submitted.

DOLINSKY, Y., and Y. KIFER (2008): Binomial Approximations of Shortfall Risk for Game Options, *Ann. Appl. Probab.* 18, 1737–1770.

DOLINSKY, Y., and Y. KIFER (2011): Binomial Approximations for Barrier Options of Israeli Style, *Ann. Dyn. Games*, XI, 447–469.

GUASONI, P. (2002a): Risk Minimization under Transaction Costs, *Finance Stoch.* 6, 91–113.

GUASONI, P. (2002b): Optimal Investment with Transaction Costs and without Semimartingales, *Ann. Appl. Probab.* 12, 1227–1246.

JAKUBENAS, P., S. LEVENTAL, and M. RYZNAR (2003): The Super-Replication Problem via Probabilistic Methods, *Ann. Appl. Probab.* 13, 742–773.

KAMIZONO, K. (2001): *Partial Hedging under Proportional Transaction Costs*. PhD thesis, Columbia University.

KAMIZONO, K. (2003): Partial Hedging under Transaction Costs, *SIAM J.Control Opt.* 5, 1545–1558.

KIFER, Y. (2000): Game Options, *Finance Stoch.* 4, 443–463.

KIFER, Y. (2006): Error Estimates for Binomial Approximiations of Game Options, *Ann. Appl. Probab.* 16, 984–1033.

LEVENTAL, S., and A. V. SKOROHOD (1997): On the Possibility of Hedging Options in the Presence of Transaction Costs, *Ann. Appl. Probab.* 7, 410–443.

MULINACCI, S. (2011): The Efficient Hedging Problem for American Options, *Finance Stoch.*, 15, 365–397.

PESKIR, G., and A. N. SHIRYAEV (2006): *Optimal Stopping and Free-Boundary Problems*. ETH Zurich: Birkhauser.

PROTTER, M. H., and C. B. MORREY (1991): *A First Course in Real Analysis*, New York: Springer-Verlag.

SONER, H. M., S. E. SHREVE, and J. CVITANIC (1995): There Is No Nontrivial Hedging Portfolio for Option Pricing with Transaction Costs, *Ann. Appl. Probab.* 5, 327–355.

TRIVELLATO, B. (2009): Replication and Shortfall Risk in a Binomial Model with Transaction Costs, *Math Meth Oper Res.* 69, 1–26.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 24, No. 3 (July 2014), 598–626

BLACK–SCHOLES REPRESENTATION FOR ASIAN OPTIONS

JAN VECER

Frankfurt School of Finance and Management

Asian options are securities with a payoff that depends on the average of the underlying stock price over a certain time interval. We identify three natural assets that appear in pricing of the Asian options, namely a stock S, a zero coupon bond B^T with maturity T, and an abstract asset A (an "average asset") that pays off a weighted average of the stock price number of units of a dollar at time T. It turns out that each of these assets has its own martingale measure, allowing us to obtain Black–Scholes type formulas for the fixed strike and the floating strike Asian options. The model independent formulas are analogous to the Black–Scholes formula for the plain vanilla options; they are expressed in terms of probabilities under the corresponding martingale measures that the Asian option will end up in the money. Computation of these probabilities is relevant for hedging. In contrast to the plain vanilla options, the probabilities for the Asian options do not admit a simple closed form solution. However, we show that it is possible to obtain the numerical values in the geometric Brownian motion model efficiently, either by solving a partial differential equation numerically, or by computing the Laplace transform. Models with stochastic volatility or pure jump models can be also priced within the Black–Scholes framework for the Asian options.

KEY WORDS: Asian option, Black–Scholes formula, hedging.

1. INTRODUCTION

An Asian option is a contract whose payoff depends on the average of the underlying stock price over a certain time interval. The Asian option payoff can be expressed in terms of an abstract "average asset" $A(t)$ that is defined by its payoff $A(T) = \int_0^T S(t)\eta(t)\,dt$. The weights are given by a measure $\eta(t)$. For instance, the fixed strike Asian option pays off $(A(T) - K \cdot B^T(T))^+$, the floating strike Asian option pays off $(A(T) - K \cdot S(T))^+$. The main contributions of this paper are the following. First, we show that the average asset A has a positive price and admits a model free replication, and thus it has its own martingale measure $\mathbb{P}^A$ (Theorem 2.1). From here we immediately get the Black–Scholes representation of the value of the Asian option in terms of the corresponding probabilities that the option will end up in the money (Theorem 2.3). This representation is model independent. In the case of the geometric Brownian motion model, it turns out that the Black–Scholes representation of the price also agrees with its hedging portfolio which is also the case with plain vanilla options (Theorem 4.2). However, in contrast to the plain vanilla options, the probabilities for the Asian options do not admit a simple closed form solution. We show in our paper that both the price and the hedge can be computed numerically from the partial differential equations (Theorem 4.1, Remark 4.3). In the case of exponential weighting, it is possible to obtain the Laplace transform representation of

Manuscript received July 2011; final revision received May 2012.
Address correspondence to Jan Vecer, Frankfurt School of Finance and Management, Sonnemannstrasse 9-11, Frankfurt 60314, Germany; e-mail: j.vecer@fs.de.

DOI: 10.1111/mafi.12012

the probabilities that appear in the Black–Scholes formulas for Asian options (Theorem 5.1). It turns out that the problem of pricing Asian options is simpler for the case of the floating strike. We show in Theorem 6.2 that under some broad conditions, it is possible to transform the pricing problem of the fixed strike option to the pricing problem of the floating strike option. The Black–Scholes approach for the Asian options can also be extended to the stochastic volatility models and to the pure jump models (Theorems 7.1 and 7.3).

Asian options were first introduced by Boyle and Emanuel (1980), and later extensively studied in the literature. For the geometric Brownian motion model, one can find the corresponding partial differential equation for the price of the Asian option. However, the exact form of the partial differential equation depends on the specific selection of the assets that enter a given contract, and on the choice of the reference asset. These choices are not unique and thus one can obtain different formulations of the corresponding pricing problem. Traditionally, the triplet of the assets that was used in the pricing of Asian options was the underlying stock S, money in terms of a dollar \$, and the asset R that represents the running average of the stock price process defined as $R(t) = \frac{1}{t}\int_0^t S(u)\,du$. The problem with this choice is twofold. One is that currencies do not have their own martingale measures, thus a better candidate for the numeraire is the corresponding bond B^T that comes with the T-forward measure $\mathbb{P}^T$. The T-forward measure $\mathbb{P}^T$ agrees with the risk neutral measure $\tilde{\mathbb{P}}$ when the interest rate is deterministic.

However, the major issue is that the running average R cannot be replicated by a self-financing trading strategy, so it does not correspond to an asset with its own martingale measure. Moreover, the price of R does not admit a martingale evolution under both the T-forward measure $\mathbb{P}^T$ and the stock measure $\mathbb{P}^S$. The first formulation of a partial differential equation in this setup appeared in Ingersoll (1987), where the T-forward measure (or equivalently a risk neutral measure in this case) was used for pricing. Rogers and Shi (1995) formulated this problem using a stock as a numeraire. Andreasen (1998) found a partial differential equation for discretely sampled Asian options, but since the price of the running average R is making discrete jumps at the sampling times, the solution of the partial differential equation also suffers from the jumps, and it has to be pasted at the sampling times. Several authors used alternative methods finding the price of the Asian options, but these results were limited to continuously sampled options with exponential weighting. For instance, Geman and Yor (1993) computed the Laplace transform of the price of continuously sampled Asian options. Linetsky (2004) found a spectral expansion of the Asian option price that involves Whittaker functions.

Vecer (2001, 2002) found a formulation that replaced the running average R with the Asian forward F that is already a self-financing object, and thus its price evolution is a martingale under both the T-forward measure $\mathbb{P}^T$ and the stock measure $\mathbb{P}^S$. This formulation leads to a simple partial differential equation when the stock is chosen as a numeraire. This pricing partial differential equation was also found independently by Hoogland and Neumann (2000) who used symmetry arguments. Fouque and Han (2003) extended this approach to stochastic volatility models, and Vecer and Xu (2004) to models with jumps. Bayraktar and Xing (2011) found an iterative numerical method for solving the corresponding partial integro-differential equation for Asian options. However, the price of the Asian forward F can become zero or negative, and thus it cannot be used as a numeraire. In contrast to the Asian forward F, the average asset A has always a positive price. It corresponds to the choice of a zero strike in the Asian forward F.

Some authors noted equivalences of Asian options with fixed and floating strikes. These results appeared in Henderson and Wojakowski (2002), and in Eberlein and

Papapantoleon (2005). Our paper naturally extends these results, it finds the most general condition for the evolution of the stock price such that the price of the fixed strike option is equal to the price of the floating strike option.

2. BLACK–SCHOLES REPRESENTATION FOR ASIAN OPTIONS

Let us first introduce notation that we use in this article. By X or Y we mean an asset rather than the price of the asset. One can think about X or Y as names of the assets that have no numerical meaning. We write $X(t)$ or $Y(t)$ in the situation when the asset is required at time t for trading, hedging, or settling a financial contract. The price of an asset is a pairwise relationship of two assets, which we denote by $X_Y(t)$: the number of assets Y required to obtain a unit of an asset X. The asset Y is known as a reference asset, or as a numeraire. We will also use the relationship

$$X_Z(t) = X_Y(t) \cdot Y_Z(t),$$

known as the change of numeraire formula. A simple consequence of the change of numeraire formula is that we can obtain the price $X_Y(t)$ from the dollar prices $X_\$(t)$ and $Y_\$(t)$ by

$$X_Y(t) = \frac{X_\$(t)}{Y_\$(t)}.$$

Asian options are contracts that depend on three assets: a stock S, a dollar \$, and the asset that represents the average of the price process $S_\$(t)$. The average price process is captured by an asset A defined by its payoff

$$A(T) = \left[\int_0^T S_\$(t)\eta(t)\,dt\right] \cdot \$(T), \tag{2.1}$$

which means that $[\int_0^T S_\$(t)\eta(t)\,dt]$ units of a dollar are delivered at time T. This is like an Asian forward with zero strike. We will call A an average asset. The weights are given by the measure η. Averaging is chosen to be continuous with uniform weights when $\eta(t) = \frac{1}{T}$, or discrete when $\eta(t) = \frac{1}{n}\sum_{k=1}^{n} \delta_{(\frac{k}{n}T)}(t)$. Since the dollar stock price $S_\$$ is not a martingale (the dollar \$ does not have its own martingale measure), it is useful to rewrite the asset A in terms of the prices that have martingale evolution under the properly chosen numeraire, which is the bond B^T in this case. We can apply the change of numeraire formula and write

$$S_\$(t) = S_{B^T}(t) \cdot B_\$^T(t).$$

Assuming that the dollar price of the bond $B_\$^T(t)$ is deterministic, and given by $B_\$^T(t) = e^{-r(T-t)}$, we have

$$S_\$(t) = e^{-r(T-t)} \cdot S_{B^T}(t).$$

Therefore the payoff of the asset A can be expressed as

$$A(T) = \left[\int_0^T S_{B^T}(t) e^{-r(T-t)}\eta(t)\,dt\right] \cdot B^T(T),$$

where the price process $S_{B^T}(t)$ is a $\mathbb{P}^T$ martingale. Thus we can rewrite the average asset A as

$$A(T) = \left[\int_0^T S_{B^T}(t)\mu(t)\,dt\right] \cdot B^T(T),$$

where $\mu(t) = e^{-r(T-t)}\eta(t)$. The weight corresponding to the uniform continuous averaging is $\mu(t) = \frac{1}{T}e^{-r(T-t)}$, and the weight corresponding to the discrete uniform averaging is $\mu(t) = \frac{1}{n}\sum_{k=1}^{n} e^{-r(T-t)}\delta_{(\frac{k}{n}T)}(t)$.

Let us first show that the asset A has its own martingale measure. When $S(t)$ and $B^T(t)$ have a positive price, the average asset $A(t)$ has also a positive price and it can be replicated by a self-financing trading strategy as proved in the following theorem. This result was noted in the earlier literature (for instance in Vecer 2002), we just mention this result in order to introduce the corresponding martingale measure $\mathbb{P}^A$ and list some of the properties of the asset A.

THEOREM 2.1. *The self-financing replicating portfolio for the average asset contract A that pays off*

$$A(T) = \left[\int_0^T S_{B^T}(t)\mu(t)\,dt\right] \cdot B^T(T) \tag{2.2}$$

is given by

$$A(t) = \left[\int_t^T \mu(s)\,ds\right] \cdot S(t) + \left[\int_0^t S_{B^T}(s)\mu(s)\,ds\right] \cdot B^T(t) \tag{2.3}$$

$$= \left[\int_t^T \mu(s)\,ds\right] \cdot S(t) + \left[\int_0^t S_{\$}(s)\eta(s)\,ds\right] \cdot B^T(t). \tag{2.4}$$

This result does not depend on the dynamics of the price $S_{B^T}(t)$.

Proof. The price $A_{B^T}(t)$ is a $\mathbb{P}^T$ martingale, and thus

$$\begin{aligned}
A_{B^T}(t) = \mathbb{E}_t^T[A_{B^T}(T)] &= \mathbb{E}_t^T\left[\int_0^T S_{B^T}(s)\mu(s)\,ds\right] \\
&= \mathbb{E}_t^T\left[\int_0^t S_{B^T}(s)\mu(s)\,ds + \int_t^T S_{B^T}(s)\mu(s)\,ds\right] \\
&= \left[\int_0^t S_{B^T}(s)\mu(s)\,ds\right] + \left[\int_t^T \mu(s)\,ds\right] \cdot S_{B^T}(t),
\end{aligned}$$

where we used the fact that the price $S_{B^T}(t)$ is a $\mathbb{P}^T$ martingale. Therefore

$$A(t) = \left[\int_t^T \mu(s)\,ds\right] \cdot S(t) + \left[\int_0^t S_{B^T}(s)\mu(s)\,ds\right] \cdot B^T(t).$$

This represents a self-financing portfolio in the assets S and B^T. To see that, note that

$$\begin{aligned}
d(A_{B^T}(t)) &= d\left(\left[\int_t^T \mu(s)\,ds\right]\cdot S_{B^T}(t) + \left[\int_0^t S_{B^T}(s)\mu(s)\,ds\right]\right) \\
&= \left[\int_t^T \mu(s)\,ds\right]\cdot d(S_{B^T}(t)) + S_{B^T}(t)\cdot d\left[\int_t^T \mu(s)\,ds\right] + d\left[\int_0^t S_{B^T}(s)\mu(s)\,ds\right] \\
&= \left[\int_t^T \mu(s)\,ds\right]\cdot d(S_{B^T}(t)). \qquad \square
\end{aligned}$$

REMARK 2.2. Theorem 2.1 cannot be extended to the situation when the bond price $B_\$^T(t)$ is stochastic, in which case the average asset A cannot be perfectly replicated in general. The stock price $S_\$\,(t)$ would depend on two sources of randomness, namely on $S_{B^T}(t)$ and on $B_\$\,(t)$, which leads to an incomplete market. Consider the simplest version of the average asset when S is the bond B, in which case

$$A(T) = \left[\int_0^T B_\$^T(t)\eta(t)\,dt\right]\cdot B^T(T).$$

Then even in the trivial case when the weighting reduces to a single time point $T_1 < T$ ($\eta(t) = \delta_{T_1}(t)$), the average asset

$$A(T) = \left[B_\$^T(T_1)\right]\cdot B^T(T)$$

cannot be perfectly replicated. For instance when $\mathbb{P}_t^T(B_\$^T(T_1) = 0.95) = \frac{1}{2}$, $\mathbb{P}_t^T(B_\$^T(T_1) = 0.85) = \frac{1}{2}$ for $t < T_1$, one cannot hedge both scenarios at the same time. Perfect hedge requires a deterministic value of $B_\$^T(T_1)$.

Notation: We will write

$$A(t) = \bar{\Delta}^S(t)\cdot S(t) + \bar{\Delta}^T(t)\cdot B^T(t)$$

in the following text, where $\bar{\Delta}^S(t) = \int_t^T \mu(s)\,ds$, and $\bar{\Delta}^T(t) = \int_0^t S_{B^T}(s)\mu(s)\,ds$. Note that

$$dA_{B^T}(t) = \bar{\Delta}^S(t)\,dS_{B^T}(t),$$

and

$$dA_S(t) = \bar{\Delta}^T(t)\,dB_S^T(t).$$

Asian options can be viewed as contracts on three underlying assets: S, B^T, and the average asset A. The most typical traded on the market are the *fixed strike Asian call option U* that pays off

$$U(T) = (A(T) - K\cdot B^T(T))^+, \tag{2.5}$$

and the *floating strike Asian call option V* that pays off

$$V(T) = (A(T) - K\cdot S(T))^+, \tag{2.6}$$

and their put option counterparts. We can rewrite the payoff of the fixed strike Asian call option as a combination of two digitals:

$$U(T) = (A(T) - K \cdot B^T(T))^+ = \mathbb{I}(A_{B^T}(T) \geq K) \cdot A(T) - K \cdot \mathbb{I}(A_{B^T}(T) \geq K) \cdot B^T(T). \tag{2.7}$$

The price of a digital option that pays off $\mathbb{I}(A_{B^T}(T) \geq K)$ units of an asset $A(T)$ at time T is $\mathbb{P}_t^A(A_{B^T}(T) \geq K)$ units of the asset $A(t)$ at time t. This follows from the First Fundamental Theorem of Asset Pricing when A is used as a numeraire. Similarly, the price of a digital option that pays off $\mathbb{I}(A_{B^T}(T) \geq K)$ units of an asset $B^T(T)$ at time T is $\mathbb{P}_t^T(A_{B^T}(T) \geq K)$ units of the asset $B^T(t)$ at time t. Combining these results, we get the following theorem.

THEOREM 2.3 [Price representation of the Asian option]. *The price of the fixed strike Asian call option at time t is given by*

$$U(t) = \mathbb{P}_t^A(A_{B^T}(T) \geq K) \cdot A(t) - K \cdot \mathbb{P}_t^T(A_{B^T}(T) \geq K) \cdot B^T(t) \tag{2.8}$$

$$\begin{aligned} &= \left[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_{B^T}(T) \geq K)\right] \cdot S(t) \\ &\quad + \left[\bar{\Delta}^T(t) \cdot \mathbb{P}_t^A(A_{B^T}(T) \geq K) - K \cdot \mathbb{P}_t^T(A_{B^T}(T) \geq K)\right] \cdot B^T(t), \end{aligned} \tag{2.9}$$

and the price of the floating strike Asian call option at time t is given by

$$V(t) = \mathbb{P}_t^A(A_S(T) \geq K) \cdot A(t) - K \cdot \mathbb{P}_t^S(A_S(T) \geq K) \cdot S(t) \tag{2.10}$$

$$\begin{aligned} &= \left[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_S(T) \geq K) - K \cdot \mathbb{P}_t^S(A_S(T) \geq K)\right] \cdot S(t) \\ &\quad + \left[\bar{\Delta}^T(t) \cdot \mathbb{P}_t^A(A_S(T) \geq K)\right] \cdot B^T(t). \end{aligned} \tag{2.11}$$

The above formulas can be viewed as Black–Scholes formulas applied to the assets A and B^T in the case of the fixed strike option (equation (2.8)), and the assets A and S in the case of the floating strike option (equation (2.10)). Since the asset A itself is a combination of a stock S and a bond B^T, we can rewrite the price of the fixed strike Asian option in terms of the stock S and the bond B^T only as in equation (2.9), and the price of the floating strike Asian option as in equation (2.11). The Black–Scholes formula turns out to be robust as shown in El Karoui, Jeanblanc-Picqué, and Shreve (1998).

Thus the problem of Asian option pricing can be reduced to computation of the probabilities under the respective numeraire measures that the option will end up in the money. Note that the formulas in Theorem 2.3 are model independent, the exact values depend on the specified evolution of the underlying prices. Even in the simplest geometric Brownian motion model, the distributions of the prices $A_{B^T}(T)$ and $A_S(T)$ are far from trivial. As we will show in the text that follows, we can compute these probabilities numerically from partial differential equations. A Laplace transform representation can be obtained in the special case of continuous weighting with exponential weights.

The advantage of the Black–Scholes representation for the Asian options is related to hedging. In the geometric Brownian motion model, it is well known that a plain vanilla European option with a payoff $(S(T) - K \cdot B^T(T))^+$ is hedged by a position of $\mathbb{P}^S(S_\$(T) \geq K)$ units of a stock S. A similar result holds for Asian options. We will show in the text that follows that the hedging position for the fixed strike Asian option is to hold $[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_{B^T}(T) \geq K)]$ units of the stock S at time t, and for the floating strike Asian option is to hold $[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_S(T) \geq K) - K \cdot \mathbb{P}_t^S(A_S(T) \geq K)]$ units of the stock S at time t.

3. GENERAL PAYOFFS AND PRICING IN MARKOV MODELS

The remaining goal is to determine the probabilities $\mathbb{P}_t^A(A_S(T) \geq K)$ and $\mathbb{P}_t^S(A_S(T) \geq K)$ that appear in the formula for the floating strike Asian option, and the probabilities $\mathbb{P}_t^A(A_{B^T}(T) \geq K)$ and $\mathbb{P}_t^T(A_{B^T}(T) \geq K)$ that appear in the formula for the fixed strike Asian option, and show their relevance to hedging. These are special cases of prices of Asian options with a particular payoff. To determine the above probabilities, let us first define the most general payoff of the Asian option.

DEFINITION 3.1. *An Asian option is a contract that pays off one of the following:*

- $f^T(S_{B^T}(T), A_{B^T}(T))$ *units of the bond* B^T,
- $f^S(A_S(T), B_S^T(T))$ *units of the stock* S,
- $f^A(S_A(T), B_A^T(T))$ *units of the average asset* A.

The payoff function depends on the choice of the reference asset. Should these payoffs represent the same contract, we must have

$$f^T(x, y) = f^S\left(\frac{y}{x}, \frac{1}{x}\right) \cdot x, \quad f^S(x, y) = f^T\left(\frac{1}{y}, \frac{x}{y}\right) \cdot y,$$

$$f^T(x, y) = f^A\left(\frac{x}{y}, \frac{1}{y}\right) \cdot y, \quad f^A(x, y) = f^T\left(\frac{x}{y}, \frac{1}{y}\right) \cdot y,$$

$$f^S(x, y) = f^A\left(\frac{1}{x}, \frac{y}{x}\right) \cdot x, \quad f^A(x, y) = f^S\left(\frac{1}{x}, \frac{y}{x}\right) \cdot x.$$

Let us show for instance the relationship of the two payoff functions f^S and f^A. In order that they represent the same contract, we must have

$$f^S\left(A_S(T), B_S^T(T)\right) \cdot S(T) = f^A\left(S_A(T), B_A^T(T)\right) \cdot A(T),$$

or in other words

$$f^S\left(A_S(T), B_S^T(T)\right) = f^A\left(S_A(T), B_A^T(T)\right) \cdot A_S(T).$$

The x variable stands for $A_S(T)$, the y variable stands for $B_S^T(T)$. It is easy to see that $S_A(T)$ is now $\frac{1}{x}$, and $B_A^T(T) = B_S^T(T) \cdot S_A(T)$ is represented by $\frac{y}{x}$.

EXAMPLE 3.2. The Asian call option with a fixed strike pays off

$$(A(T) - K \cdot B^T(T))^+.$$

The payoff can be settled in three equivalent ways:

$$(A_{B^T}(T) - K)^+ \cdot B^T(T) = \left(A_S(T) - K \cdot B_S^T(T)\right)^+ \cdot S(T) = \left(1 - K \cdot B_A^T(T)\right)^+ \cdot A(T).$$

This corresponds to the payoff functions $f^T(x, y) = (y - K)^+$, $f^S(x, y) = (x - K \cdot y)^+$, or $f^A(x, y) = (1 - K \cdot y)^+$ in the above definition of the Asian option.

The Asian call option with a floating strike pays off

$$(A(T) - K \cdot S(T))^+ .$$

The payoff can be settled in the following three ways:

$$(A_{B^T}(T) - K \cdot S_{B^T}(T))^+ \cdot B^T(T) = (A_S(T) - K)^+ \cdot S(T) = (1 - K \cdot S_A(T))^+ \cdot A(T),$$

which corresponds to the payoff functions $f^T(x, y) = (y - K \cdot x)^+$, $f^S(x, y) = (x - K)^+$, or $f^A(x, y) = (1 - K \cdot x)^+$.

Asian options with fixed or floating strike are the two most typical Asian option contracts. Let V denote an Asian option contract. Then we can express its price using the numeraires B^T, S, and A as

$$V(t) = \mathbb{E}_t^T[V_{B^T}(T)] \cdot B^T(t) = \mathbb{E}_t^S[V_S(T)] \cdot S(t) = \mathbb{E}_t^A[V_A(T)] \cdot A(t).$$

When the pricing problem is Markovian in the prices of the underlying assets, we can introduce the following functions:

$$\begin{aligned} u^T(t, x, y) &= \mathbb{E}^T\left[V_{B^T}(T) \mid S_{B^T}(t) = x,\ A_{B^T}(t) = y\right] \\ &= \mathbb{E}^T\left[f^T\left(S_{B^T}(T),\ A_{B^T}(T)\right) \mid S_{B^T}(t) = x,\ A_{B^T}(t) = y\right], \end{aligned}$$

$$\begin{aligned} u^S(t, x, y) &= \mathbb{E}^S\left[V_S(T) \mid A_S(t) = x,\ B_S^T(t) = y\right] \\ &= \mathbb{E}^S\left[f^S\left(A_S(T),\ B_S^T(T)\right) \mid A_S(t) = x,\ B_S^T(t) = y\right], \end{aligned}$$

$$\begin{aligned} u^A(t, x, y) &= \mathbb{E}^A\left[V_A(T) \mid S_A(t) = x,\ B_A^T(t) = y\right] \\ &= \mathbb{E}^A\left[f^A\left(S_A(T),\ B_A^T(T)\right) \mid S_A(t) = x,\ B_A^T(t) = y\right]. \end{aligned}$$

Thus we can write

$$\begin{aligned} V(t) &= u^T\left(t,\ S_{B^T}(t),\ A_{B^T}(t)\right) \cdot B^T(t) = u^S\left(t,\ A_S(t),\ B_S^T(t)\right) \cdot S(t) \\ &= u^A\left(t,\ S_A(t),\ B_A^T(t)\right) \cdot A(t), \end{aligned}$$

giving us the following relationships between u^T, u^S, and u^A:

$$u^T(t, x, y) = u^S\left(t, \frac{y}{x}, \frac{1}{x}\right) \cdot x, \quad u^S(t, x, y) = u^T\left(t, \frac{1}{y}, \frac{x}{y}\right) \cdot y, \tag{3.1}$$

$$u^T(t, x, y) = u^A\left(t, \frac{x}{y}, \frac{1}{y}\right) \cdot y, \quad u^A(t, x, y) = u^T\left(t, \frac{x}{y}, \frac{1}{y}\right) \cdot y, \tag{3.2}$$

$$u^S(t, x, y) = u^A\left(t, \frac{1}{x}, \frac{y}{x}\right) \cdot x, \quad u^A(t, x, y) = u^S\left(t, \frac{1}{x}, \frac{y}{x}\right) \cdot x. \tag{3.3}$$

4. PRICING AND HEDGING IN GEOMETRIC BROWNIAN MOTION MODEL

The prices of assets should be martingales under their corresponding numeraire measures. Since we have three underlying assets B^T, S, and A, we have six price processes to consider: $S_{B^T}(t)$, $A_{B^T}(t)$, $B^T_S(t)$, $A_S(t)$, $S_A(t)$, and $B^T_A(t)$. The price processes $S_{B^T}(t)$, and $A_{B^T}(t)$ are $\mathbb{P}^T$ martingales, the price processes $B^T_S(t)$, and $A_S(t)$ are $\mathbb{P}^S$ martingales, and the price processes $S_A(t)$, and $B^T_A(t)$ are $\mathbb{P}^A$ martingales.

In the geometric Brownian motion model we assume the following price dynamics:

$$dS_{B^T}(t) = \sigma\, S_{B^T}(t)\, dW^T(t), \tag{4.1}$$

which already implies

$$d\,B^T_S(t) = \sigma\, B^T_S(t)\, dW^S(t). \tag{4.2}$$

The two Brownian motions W^T and W^S are perfectly negatively correlated, so

$$dW^S(t)\cdot dW^T(t) = -dt.$$

The evolution of $A_{B^T}(t)$ follows from the hedging formula for the average asset A:

$$dA_{B^T}(t) = \bar{\Delta}^S(t)\, dS_{B^T}(t) = \sigma\,\bar{\Delta}^S(t) S_{B^T}(t)\, dW^T(t). \tag{4.3}$$

Note that this evolution is not Markovian in $A_{B^T}(t)$ since it depends on another process $S_{B^T}(t)$, but it is Markovian in the pair $(S_{B^T}(t),\ A_{B^T}(t))$. Thus even when the Asian option contract payoff depends only on the assets A and B^T, such as in the case of the fixed strike Asian option, the pricing partial differential equation that corresponds to the pricing measure $\mathbb{P}^T$ would depend on both prices $S_{B^T}(t)$ and $A_{B^T}(t)$.

The evolution of the $A_S(t)$ price can be rewritten as

$$\begin{aligned} dA_S(t) &= \bar{\Delta}^T(t) dB^T_S(t) \\ &= [A_{B^T}(t) - \bar{\Delta}^S(t)\cdot S_{B^T}(t)]\, dB^T_S(t) \\ &= [A_{B^T}(t) - \bar{\Delta}^S(t)\cdot S_{B^T}(t)]\,\sigma\, B^T_S(t)\, dW^S(t) \\ &= \sigma[A_S(t) - \bar{\Delta}^S(t)]\, dW^S(t). \end{aligned} \tag{4.4}$$

The second equality $\bar{\Delta}^T(t) = A_{B^T}(t) - \bar{\Delta}^S_t \cdot S_{B^T}(t)$ follows from the relation $A(t) = \bar{\Delta}^S(t)\cdot S(t) + \bar{\Delta}^T(t)\cdot B^T(t)$. The reason for writing the evolution of $A_S(t)$ in terms of $\bar{\Delta}^S(t)$ rather than in terms of $\bar{\Delta}^T(t)$ is that $\bar{\Delta}^S(t)$ is deterministic, while $\bar{\Delta}^T(t)$ is stochastic. This means that unlike the price evolution of $A_{B^T}(t)$, the price evolution of $A_S(t)$ is Markovian in just one variable, and thus contracts whose payoff depends only on the assets A and S admit a simpler partial differential equation with only one spatial variable.

Let us determine the evolution of the remaining prices: $B^T_A(t)$, and $S_A(t)$. From Itô's formula we have

$$\begin{aligned} dB^T_A(t) = d[A_{B^T}(t)]^{-1} &= -[A_{B^T}(t)]^{-2} dA_{B^T}(t) + [A_{B^T}(t)]^{-3} d^2 A_{B^T}(t) \\ &= -[B^T_A(t)]^2 \sigma\,\bar{\Delta}^S(t) S_{B^T}(t)\, dW^T(t) + [B^T_A(t)]^3 \sigma^2 [\bar{\Delta}^S(t)]^2 [S_{B^T}(t)]^2\, dt \\ &= \sigma\,\bar{\Delta}^S(t) B^T_A(t) S_A(t)[-dW^T(t) + \sigma\,\bar{\Delta}^S(t) S_A(t)\, dt]. \end{aligned}$$

According to the First Fundamental Theorem of Asset Pricing, the evolution of $B_A^T(t)$ has to be a martingale under the corresponding $\mathbb{P}^A$ measure. Thus we have

$$dB_A^T(t) = \sigma\,\bar{\Delta}^S(t)\,B_A^T(t)S_A(t)\,dW^A(t), \tag{4.5}$$

where

$$dW^A(t) = -dW^T(t) + \sigma\,\bar{\Delta}^S(t)S_A(t)\,dt \tag{4.6}$$

is a Brownian motion under the $\mathbb{P}^A$ measure. Note that $W^A(t)$ is perfectly negatively correlated with $W^T(t)$, and it is perfectly correlated with $W^S(t)$. The evolution of $S_A(t)$ is given by

$$\begin{aligned}
dS_A(t) = d[A_S(t)]^{-1} &= -[A_S(t)]^{-2}dA_S(t) + [A_S(t)]^{-3}d^2A_S(t)\\
&= -[S_A(t)]^2\sigma[A_S(t) - \bar{\Delta}^S(t)]\,dW^S(t) + [S_A(t)]^3\sigma^2[A_S(t) - \bar{\Delta}^S(t)]^2\,dt\\
&= \sigma\,S_A(t)[\bar{\Delta}^S(t)S_A(t) - 1][dW^S(t) - \sigma[1 - \bar{\Delta}^S(t)S_A(t)]\,dt].
\end{aligned}$$

Therefore

$$dS_A(t) = \sigma\,S_A(t)[\bar{\Delta}^S(t)S_A(t) - 1]dW^A(t), \tag{4.7}$$

which is a martingale under the $\mathbb{P}^A$ measure.

Recall that Asian options pay off either $f^T(S_{B^T}(T), A_{B^T}(T))$ units of $B^T(T)$, $f^S(A_S(T), B_S^T(T))$ units of $S(T)$, or $f^A(S_A(T), B_A^T(T))$ units of $A(T)$. When $f^T(x, y) = f^S(\frac{y}{x}, \frac{1}{x}) \cdot x = f^A(\frac{x}{y}, \frac{1}{y}) \cdot y$, the three payoffs represent the same contract. The price of the Asian option is determined in the next theorem.

THEOREM 4.1. *The function* $u^T(t, x, y) = \mathbb{E}^T[V_{B^T}(T) \mid S_{B^T}(t) = x,\ A_{B^T}(t) = y]$ *satisfies partial differential equation*

$$u_t^T(t, x, y) + \frac{1}{2}\sigma^2x^2\left[u_{xx}^T(t, x, y) + 2\bar{\Delta}^S(t)u_{xy}^T(t, x, y) + [\bar{\Delta}^S(t)]^2u_{yy}^T(t, x, y)\right] = 0 \tag{4.8}$$

with the terminal condition

$$u^T(T, x, y) = f^T(x, y), \tag{4.9}$$

the function $u^S(t, x, y) = \mathbb{E}^S[V_S(T) \mid A_S(t) = x,\ B_S^T(t) = y]$ *satisfies partial differential equation*

$$\begin{aligned}
u_t^S(t, x, y) + \frac{1}{2}\sigma^2\big[&(x - \bar{\Delta}^S(t))^2u_{xx}^S(t, x, y) + 2y(x - \bar{\Delta}^S(t))u_{xy}^S(t, x, y)\\
&+ y^2u_{yy}^S(t, x, y)\big] = 0,
\end{aligned} \tag{4.10}$$

with the terminal condition

$$u^S(T, x, y) = f^S(x, y), \tag{4.11}$$

and the function $u^A(t, x, y) = \mathbb{E}^A[V_A(T) \mid S_A(t) = x,\ B_A^T(t) = y]$ *satisfies partial differential equation*

$$u_t^A(t, x, y) + \frac{1}{2}\sigma^2 x^2\big([x\bar{\Delta}^S(t) - 1]^2 \cdot u_{xx}^A(t, x, y) + 2y[x\bar{\Delta}^S(t) - 1] \cdot u_{xy}^A(t, x, y) + y^2[\bar{\Delta}^S(t)]^2 \cdot u_{yy}^A(t, x, y)\big) = 0, \tag{4.12}$$

with the terminal condition

$$u^A(T, x, y) = f^A(x, y). \tag{4.13}$$

Proof. Heuristically it follows directly from the Feynman-Kac Theorem. However, the standard version of the Feynman-Kac Theorem requires the partial differential equation to be uniformly parabolic (see Friedman 1975, chapter 6), a condition that is not satisfied in this particular case. For this degenerate type of equations, the notion of viscosity solutions needs to be employed to ensure the expectation (value function) to be a solution of the associated PDE.[1] □

The pricing PDEs for Asian options listed above are related to the ones used in the existing literature as they describe the same pricing problem. However, the choice of assets A, S, and B^T makes the underlying prices martingales under the respective numeraire measures, and thus the PDEs in Theorem 4.1 are given in their simplest form. In particular, they do not have any first-order spatial derivatives. The previous literature also did not use the average asset A as a numeraire, so the PDE for u^A is novel. Theorem 4.1 is useful in the proof of the following theorem.

THEOREM 4.2 [Hedging representation of the Asian option]. *In the geometric Brownian motion model, the hedge for the fixed strike Asian option agrees with*

$$\begin{aligned} P(t) &= \mathbb{P}_t^A(A_{B^T}(T) \geq K) \cdot A(t) - K \cdot \mathbb{P}_t^T(A_{B^T}(T) \geq K) \cdot B^T(t) \\ &= \big[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_{B^T}(T) \geq K)\big] \cdot S(t) \\ &\quad + \big[\bar{\Delta}^T(t) \cdot \mathbb{P}_t^A(A_{B^T}(T) \geq K) - K \cdot \mathbb{P}_t^T(A_{B^T}(T) \geq K)\big] \cdot B^T(t), \end{aligned} \tag{4.14}$$

meaning that one should hold $[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_{B^T}(T) \geq K)]$ *units of the stock* S *at time* t. *Similarly, the hedge for the floating strike Asian option agrees with*

$$\begin{aligned} P(t) &= \mathbb{P}_t^A(A_S(T) \geq K) \cdot A(t) - K \cdot \mathbb{P}_t^S(A_S(T) \geq K) \cdot S(t) \\ &= \big[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_S(T) \geq K) - K \cdot \mathbb{P}_t^S(A_S(T) \geq K)\big] \cdot S(t) \\ &\quad + \big[\bar{\Delta}^T(t) \cdot \mathbb{P}_t^A(A_S(T) \geq K)\big] \cdot B^T(t), \end{aligned} \tag{4.15}$$

meaning that one should hold $[\bar{\Delta}^S(t) \cdot \mathbb{P}_t^A(A_S(T) \geq K) - K \cdot \mathbb{P}_t^S(A_S(T) \geq K)]$ *units of the stock* S *at time* t.

Proof. Let us start with the floating strike case. Define $u^S(t, x) = \mathbb{E}^S[(A_S(T) - K)^+ \mid A_S(t) = x]$. Then $u^S(t, A_S(t))$ is a $\mathbb{P}^S$ martingale and the hedging portfolio $P(t)$

[1] The author would like to thank the referee for pointing out this fact.

that is worth $u^S(t, A_S(t))$ units of a stock $S(t)$ satisfies in the geometric Brownian motion model

$$dP_S(t) = du^S(t, A_S(t)) = u_x^S(t, A_S(t))dA_S(t). \tag{4.16}$$

But we have that

$$\begin{aligned} u_x^S(t, x) &= \frac{d}{dx}\mathbb{E}_t^S(A_S(T) - K)^+ \\ &= \frac{d}{dx}\mathbb{E}^S\left[\left(x \cdot \frac{A_S(T)}{A_S(t)} - K\right)^+ \,\middle|\, A_S(t) = x\right] \\ &= \frac{d}{dx}\mathbb{E}^A\left[\left(x - K \cdot \frac{S_A(T)}{S_A(t)}\right)^+ \,\middle|\, A_S(t) = x\right] \\ &= \mathbb{P}_t^A(A_S(T) \geq K). \end{aligned}$$

Thus one should hold $\mathbb{P}_t^A(A_S(T) \geq K)$ units of an asset A and $-K \cdot \mathbb{P}_t^S(A_S(T) \geq K)$ units of an asset S. Since the asset A consists of $\bar{\Delta}^S(t)$ units of a stock S and $\bar{\Delta}^T(t)$ units of a bond B^T, we get the stated result for the floating strike Asian option.

For the fixed strike case, define $u^S(t, x, y) = \mathbb{E}^S[(A_S(T) - K \cdot B_S^T(T))^+ \mid A_S(t) = x, B_S^T(t) = y]$. The pricing problem is two dimensional in space. The hedging portfolio takes the following form

$$\begin{aligned} P(t) &= [u_x^S] \cdot A(t) + [u_y^S] \cdot B^T(t) + [u^S - u_x^S \cdot A_S - u_y^S \cdot B_S^T] \cdot S(t) \\ &= [u^S + u_x^S \cdot (\bar{\Delta}^S - A_S) - u_y^S \cdot B_S^T] \cdot S(t) + [\bar{\Delta}^T \cdot u_x^S + u_y^S] \cdot B^T(t), \end{aligned}$$

meaning that one should hold $[u^S + u_x^S \cdot (\bar{\Delta}^S - A_S) - u_y^S \cdot B_S^T]$ units of the stock $S(t)$ to hedge this contract. Similarly to the floating strike Asian option, we get

$$u_x^S(t, A_S(t), B_S^T(t)) = \mathbb{P}_t^A(A_{B^T}(T) \geq K),$$

and

$$u_y^S(t, A_S(t), B_S^T(t)) = -K \cdot \mathbb{P}_t^T(A_{B^T}(T) \geq K).$$

Thus the hedging position $[u^S + u_x^S \cdot (\bar{\Delta}^S - A_S) - u_y^S \cdot B_S^T]$ simplifies to holding $\bar{\Delta}^S(t) \cdot \mathbb{P}^A(A_{B^T}(T) \geq K)$ units of the stock $S(t)$. □

REMARK 4.3. When the payoff of the Asian option depends on the assets A and S only, such as in the case of the floating strike Asian option, the equation (4.10) simplifies to

$$u_t^S(t, x) + \frac{1}{2}\sigma^2(x - \bar{\Delta}^S(t))^2 u_{xx}^S(t, x) = 0 \tag{4.17}$$

with the terminal condition $f^S(x)$, and the equation (4.12) simplifies to

$$u_t^A(t, x) + \frac{1}{2}\sigma^2 x^2(x\bar{\Delta}^S(t) - 1)^2 \cdot u_{xx}^A(t, x, y) = 0, \tag{4.18}$$

with the terminal condition $f^A(x)$. In particular, a contract with the payoff $\mathbb{I}(A_S(T) \geq K) \cdot S(T)$ that corresponds to the payoff function $f^S(x) = \mathbb{I}(x \geq K)$ (digital payoff) is worth $u^S(t, x) = \mathbb{P}^S(A_S(T) \geq K \mid A_S(t) = x)$ units of the stock $S(t)$. Thus this probability can be computed numerically by solving equation (4.17) using the corresponding terminal condition $f^S(x) = \mathbb{I}(x \geq K)$. Since $u^A(t, S_A(t)) \cdot A_S(t) = u^S(t, A_S(t))$, we can also compute $\mathbb{P}^S(A_S(T) \geq K \mid S_A(t)) = u^A(t, S_A(t)) \cdot A_S(t)$ by solving equation (4.18) for u^A using the terminal condition $f^A(x) = x \cdot \mathbb{I}(\frac{1}{K} \geq x)$. Similarly, a contract with the payoff $\mathbb{I}(A_S(T) \geq K) \cdot A(T)$ that corresponds to the payoff function $f^A(x) = \mathbb{I}(\frac{1}{K} \geq x)$ is worth $u^A(t, x) = \mathbb{P}^A(A_S(T) \geq K \mid S_A(t) = x)$ units of the average asset $A(t)$. Thus this probability can be computed numerically by solving equation (4.18) using the corresponding terminal condition $f^A(x) = \mathbb{I}(\frac{1}{K} \geq x)$. We can also compute $\mathbb{P}^A(A_S(T) \geq K \mid A_S(t)) = u^S(t, A_S(t)) \cdot S_A(t)$ by solving equation (4.17) for u^S using the terminal condition $f^S(x) = x \cdot \mathbb{I}(x \geq K)$.

Equations (4.17) and (4.18) represent an important simplification of the pricing problem for the Asian options to one spatial dimension. Although the Asian option pricing problem is a three-asset problem involving the average asset A, the stock S, and the bond B^T, the bond becomes redundant in the payoffs that involve only the assets A and S. A similar simplification does not happen when the payoff involves only the assets A and B^T, the stock S still appears implicitly as a part of the average asset A. However, under some assumptions on the dynamics of the prices, there is a connection of the fixed and the floating strike Asian options, and a reduction of the pricing problem to one spatial dimension is possible as we will show in Section 6.

Equation (4.17) already appeared in Vecer (2002) when the pricing problem involved the Asian forward F and the reference asset was chosen to be a stock S. The equation (4.18) is novel, and it is numerically equally good for Asian option pricing. Note that these partial differential equations apply for a general weighting factor $\mu(t)$, so they can be used both for the discrete and continuous weighting. The weighting also does not need to be exponential.

The method in Theorem 4.2 requires computing two values that correspond to probabilities that the option will end up in the money under the respective numeraire measures, so it is not numerically superior to the direct computation of the Asian option price that requires only one value. However, determination of the price and the hedge requires at least two values, and Theorem 4.2 gives the representation of the hedge that can be computed directly in contrast to determination of the hedge from the sensitivities of the option prices. Direct computation of the hedge from the partial differential equation has a higher numerical precision than computing it from the option sensitivities as the numerical differentiation leads to a loss of one order of the space precision. Moreover, when the weighting is chosen to be continuous and exponential, it is possible to determine the Laplace transform for the probabilities in question as we show in the following section.

5. LAPLACE TRANSFORM FOR $\mathbb{P}_t^S(A_S(T) \geq K)$ AND $\mathbb{P}_t^A(A_S(T) \geq K)$

This section considers continuous averaging when $\mu(t)\, dt = \frac{1}{T} e^{-r(T-t)}\, dt$, in which case

$$\bar{\Delta}^S(t) = \int_t^T \mu(s)\, ds = \int_t^T \frac{1}{T} e^{-r(T-s)} ds = \frac{1}{rT}(1 - e^{-r(T-t)}).$$

We determine the Laplace transform for $\mathbb{P}_t^S(A_S(T) \geq K)$ and $\mathbb{P}_t^A(A_S(T) \geq K)$ using Whittaker functions $M_{\kappa,\mu}(x)$ and $W_{\kappa,\mu}(x)$.

THEOREM 5.1. *The probability* $\mathbb{P}_t^S(A_S(T) \geq K)$ *is given by*

$$\mathbb{P}_t^S(A_S(T) \geq K) = \frac{1}{2\pi i} \int_{\rho - i\infty}^{\rho + i\infty} L^S(z, \lambda) e^{\lambda s} \, d\lambda, \tag{5.1}$$

where $s = \sigma^2(T - t)$, $\quad z = \frac{2e^{-r(T-t)}}{\sigma^2 T(A_S(t) - \bar{\Delta}^S(t))} = \frac{2e^{-r(T-t)} \cdot S_{BT}(t)}{\sigma^2 T \bar{\Delta}^T(t)} = \frac{2 S_{\$}(t)}{\sigma^2 T \int_0^t S_{\$}(s)\eta(s)\, ds}$,

$$L^S(z, \lambda) = \begin{cases} 2e^{\frac{z-a}{2}} z^{-(\kappa+1)} a^\kappa \cdot \dfrac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right) \cdot M_{\kappa,\mu}(a) \cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2} + \kappa + \mu\right) \cdot \Gamma(1 + 2\mu)}, & z \geq a, \\[2ex] 2 \cdot \dfrac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1 + 2\mu)} \cdot \left[\dfrac{M_{\kappa,\mu}(z) \cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2} + \kappa + \mu\right) \cdot z} + \dfrac{M_{\kappa+1,\mu}(z) \cdot W_{\kappa,\mu}(z)}{z} \right. & \\[2ex] \qquad \left. - e^{\frac{z-a}{2}} z^{-(\kappa+1)} a^\kappa M_{\kappa+1,\mu}(z) \cdot W_{\kappa,\mu}(a) \right], & z < a, \end{cases} \tag{5.2}$$

with $a = \frac{2}{\sigma^2 KT}$, $\kappa = \frac{r}{\sigma^2}$, $\mu = \sqrt{(\kappa + \frac{1}{2})^2 + 2\lambda}$. *The constant* ρ *in equation (5.1) is an arbitrarily constant chosen so that the contour of integration lies to the right of all singularities of the integrand.*

When $t = 0$, *the Laplace transform does not depend on the variable* z, *and it simplifies to*

$$L^S(\lambda) = 2e^{-\frac{a}{2}} a^\kappa \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right) \cdot M_{\kappa,\mu}(a)}{\left(\frac{1}{2} + \kappa + \mu\right) \cdot \Gamma(1 + 2\mu)}. \tag{5.3}$$

The probability $\mathbb{P}_t^A(A_S(T) \geq K)$ *is given by*

$$\mathbb{P}_t^A(A_S(T) \geq K) = \frac{1}{2\pi i} \int_{\rho - i\infty}^{\rho + i\infty} L^A(z, \lambda) e^{\lambda s} \, d\lambda, \tag{5.4}$$

where

$$(5.5)\quad L^A(z,\lambda) = \begin{cases} 2KS_A(t)\cdot e^{\frac{z-a}{2}} z^{-(\kappa+1)} a^{\kappa} \cdot \dfrac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)\cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2}+\kappa+\mu\right)\cdot\Gamma(1+2\mu)} \\ \qquad \cdot \left[M_{\kappa,\mu}(a) + \dfrac{M_{\kappa-1,\mu}(a)}{-\frac{1}{2}+\kappa+\mu}\right], \quad z \geq a \\ \dfrac{2aKS_A(t)\cdot\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)\cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2}+\kappa+\mu\right)\Gamma(1+2\mu)\cdot z^2} \\ \qquad \cdot \left[M_{\kappa,\mu}(z) + \dfrac{M_{\kappa-1,\mu}(z)}{-\frac{1}{2}+\kappa+\mu}\right] \\ \qquad + \dfrac{2KS_A(t)\cdot\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)}\cdot M_{\kappa+1,\mu}(z) \\ \qquad \cdot \Big[e^{\frac{z-a}{2}} a^{\kappa} z^{-(\kappa+1)} (W_{\kappa-1,\mu}(a) - W_{\kappa,\mu}(a)) \\ \qquad - \dfrac{a}{z^2}(W_{\kappa-1,\mu}(z) - W_{\kappa,\mu}(z))\Big], \quad z < a. \end{cases}$$

When $t = 0$, *we have*

$$(5.6)\quad L^A(\lambda) = 2KS_A(0)e^{-\frac{a}{2}}a^{\kappa}\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\left(\frac{1}{2}+\kappa+\mu\right)\cdot\Gamma(1+2\mu)}\cdot\left[M_{\kappa,\mu}(a) + \frac{M_{\kappa-1,\mu}(a)}{-\frac{1}{2}+\kappa+\mu}\right].$$

Proof. See Appendix.

REMARK 5.2. When $r = 0$, the Laplace transforms simplify to

$$(5.7)\qquad L^S(z,\lambda) = e^{\frac{z-a}{2}}\cdot\frac{\sqrt{\pi a}}{z\lambda}\cdot W_{1,\mu}(z)\cdot I_{\mu}\left(\frac{a}{2}\right),$$

and

$$(5.8)\qquad L^A(z,\lambda) = 2KS_A(t)e^{\frac{z-a}{2}}\cdot\frac{\Gamma\left(-\frac{1}{2}+\mu\right)\cdot W_{1,\mu}(z)}{z\cdot\left(\frac{1}{2}+\mu\right)\cdot\Gamma(1+2\mu)}\cdot\left[M_{0,\mu}(a) + \frac{M_{-1,\mu}(a)}{-\frac{1}{2}+\mu}\right].$$

The function $I_\mu(x)$ is a modified Bessel function. When $t = 0$ and $r = 0$, we get

$$L^S(\lambda) = e^{-\frac{a}{2}}\sqrt{\pi a}\cdot\frac{I_{\sqrt{2\lambda+1/4}}\left(\frac{a}{2}\right)}{\lambda}, \tag{5.9}$$

and

$$L^A(\lambda) = K S_A(0) e^{-\frac{a}{2}}\sqrt{\pi a}\cdot\frac{\left(\frac{a+1}{2}+2\lambda+\mu\right)\cdot I_\mu\left(\frac{a}{2}\right)+\frac{a}{2}\cdot I_{\mu+1}\left(\frac{a}{2}\right)}{2\lambda^2}. \tag{5.10}$$

REMARK 5.3. Given the relationship $U(t) = \mathbb{P}_t^A(A_S(T) \ge K)\cdot A(t) - K\cdot\mathbb{P}_t^S(A_S(T) \ge K)\cdot S(t)$ for the Asian option with the floating strike, the formula for $U(t)$ simplifies to

$$U(t) = \begin{cases} \left[\dfrac{Ke^{\frac{z-a}{2}}z^{-(\kappa+1)}a^\kappa}{\pi i}\displaystyle\int_{\rho-i\infty}^{\rho+i\infty}\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)\cdot M_{\kappa-1,\mu}(a)\cdot W_{\kappa+1,\mu}(z)e^{\lambda s}}{\left(-\frac{1}{2}+\kappa+\mu\right)\left(\frac{1}{2}+\kappa+\mu\right)\cdot\Gamma(1+2\mu)}d\lambda\right]\cdot S(t), & z \ge a,\\[2ex] A(t) - KS(t) \\ \quad + \left[\dfrac{Ke^{\frac{z-a}{2}}z^{-(\kappa+1)}a^\kappa}{\pi i}\displaystyle\int_{\rho-i\infty}^{\rho+i\infty}\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)\cdot M_{\kappa+1,\mu}(z)\cdot W_{\kappa-1,\mu}(a)e^{\lambda s}}{\Gamma(1+2\mu)}d\lambda\right]\cdot S(t), & z < a. \end{cases}$$

In particular, when $t = 0$, we get

$$U(0) = \left[K\cdot\frac{e^{-\frac{a}{2}}a^\kappa}{\pi i}\int_{\rho-i\infty}^{\rho+i\infty}\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)\cdot M_{\kappa-1,\mu}(a)\cdot\exp\left(\frac{2\lambda}{aK}\right)}{\left(-\frac{1}{2}+\kappa+\mu\right)\cdot\left(\frac{1}{2}+\kappa+\mu\right)\cdot\Gamma(1+2\mu)}d\lambda\right]\cdot S(0).$$

This confirms results obtained by Hoogland and Neumann (2000).

6. COMPUTING $\mathbb{P}_t^T(A_{B^T}(T) \ge K)$ AND $\mathbb{P}_t^A(A_{B^T}(T) \ge K)$

This section shows that the computation of $\mathbb{P}_t^A(A_{B^T}(T) \ge K)$ and $\mathbb{P}_t^T(A_{B^T}(T) \ge K)$ can be transformed to the problem of the computation of $\mathbb{P}_t^A(A_S(T) \ge K)$ and $\mathbb{P}_t^S(A_S(T) \ge K)$ under a relatively broad assumption. Let us first show that a seasoned (in-progress) fixed strike Asian option can be expressed as an unseasoned fixed strike Asian option with a modified strike. Thus the problem of computing $\mathbb{P}_t^A(A_{B^T}(T) \ge K)$ and $\mathbb{P}_t^T(A_{B^T}(T) \ge K)$ can be reset to time $t = 0$ using a modified strike. Theorem 6.2 shows the relationship between $\mathbb{P}^A(A_{B^T}(T) \ge K)$ and $\mathbb{P}^A(A_S(T) \ge K)$, and between $\mathbb{P}^T(A_{B^T}(T) \ge K)$ and $\mathbb{P}^S(A_S(T) \ge K)$.

LEMMA 6.1. *A fixed strike Asian option at time t with the payoff*

$$\left[\int_0^T S_{B^T}(s)\mu(s)\,ds - K\right]^+\cdot B^T(T)$$

is the same as a fixed strike Asian option at time 0 *with a payoff*

$$\left[\int_t^T S_{B^T}(s)\mu(s)\,ds - \left(K - \int_0^t S_{B^T}(s)\mu(s)\,ds\right)\right]^+ \cdot B^T(T).$$

Proof.

$$\begin{aligned}\{A_{B^T}(T) \geq K\} &\Longleftrightarrow \{A_{B^T}(T) - A_{B^T}(t) \geq K - A_{B^T}(t)\}\\ &\Longleftrightarrow \left\{A_{B^T}(T) - \int_0^t S_{B^T}(s)\mu(s)\,ds \geq K - \int_0^t S_{B^T}(s)\mu(s)\,ds\right\}\\ &\Longleftrightarrow \left\{\int_t^T S_{B^T}(s)\mu(s)\,ds \geq K - \int_0^t S_{B^T}(s)\mu(s)\,ds\right\}.\end{aligned}$$ □

THEOREM 6.2. *Let*

$$\tilde{A}(T) = \left[\int_0^T S_{B^T}(t)\mu(T-t)\,dt\right] \cdot B^T(T). \tag{6.1}$$

Then under the assumption

$$\mathcal{L}^T(A_{B^T}(T)) = \mathcal{L}^S(\tilde{A}_S(T) \cdot S_{B^T}(0)), \tag{6.2}$$

where $\mathcal{L}^Y(X)$ denotes the $\mathbb{P}^Y$ law of the random variable X, we have

$$\mathbb{P}^T(A_{B^T}(T) \geq K) = \mathbb{P}^S\left(\tilde{A}_S(T) \geq \frac{K}{S_{B^T}(0)}\right) = \mathbb{P}^S\left(\tilde{A}_S(T) \geq \frac{e^{-rT} \cdot K}{S_\$(0)}\right), \tag{6.3}$$

and

$$\mathbb{P}^A(A_{B^T}(T) \geq K) = \mathbb{P}^{\tilde{A}}\left(\tilde{A}_S(T) \geq \frac{K}{S_{B^T}(0)}\right) = \mathbb{P}^{\tilde{A}}\left(\tilde{A}_S(T) \geq \frac{e^{-rT} \cdot K}{S_\$(0)}\right). \tag{6.4}$$

Proof. Equation (6.3) follows directly from the assumption (6.2). In order to prove equation (6.4), we show first that the assumption (6.2) implies

$$\mathcal{L}^A(A_{B^T}(T)) = \mathcal{L}^{\tilde{A}}\left(\tilde{A}_S(T) \cdot S_{B^T}(0)\right).$$

$$
\begin{aligned}
\mathbb{P}^A(A_{B^T}(T) \geq K) &= \mathbb{E}^A[\mathbb{I}(A_{B^T}(T) \geq K)] \\
&= \mathbb{E}^T\left[\mathbb{I}(A_{B^T}(T) \geq K) \cdot \frac{A_{B^T}(T)}{A_{B^T}(0)}\right] \\
&= \mathbb{E}^S\left[\mathbb{I}(\tilde{A}_S(T) \cdot S_{B^T}(0) \geq K) \cdot \frac{\tilde{A}_S(T) \cdot S_{B^T}(0)}{A_{B^T}(0)}\right] \\
&= \mathbb{E}^{\tilde{A}}\left[\mathbb{I}(\tilde{A}_S(T) \cdot S_{B^T}(0) \geq K) \cdot \frac{\tilde{A}_S(T) \cdot S_{B^T}(0)}{A_{B^T}(0)} \cdot \frac{S_{\tilde{A}}(T)}{S_{\tilde{A}}(0)}\right] \\
&= \mathbb{E}^{\tilde{A}}[\mathbb{I}(\tilde{A}_S(T) \cdot S_{B^T}(0) \geq K)] \cdot \frac{\tilde{A}_S(0)}{A_S(0)} \\
&= \mathbb{P}^{\tilde{A}}\left(\tilde{A}_S(T) \geq \frac{K}{S_{B^T}(0)}\right).
\end{aligned}
$$

The second and the fourth equality follows from the change of measure via the Radon–Nikodým derivative. The ratio of $\tilde{A}_S(0)$ and $A_S(0)$ is equal to one since

$$\tilde{A}_S(0) = \int_0^T \mu(T-t)\,dt, \qquad \text{and} \qquad A_S(0) = \int_0^T \mu(t)\,dt. \qquad \square$$

The problem of finding $\mathbb{P}^T(A_{B^T}(T) \geq K)$ can be transformed to the problem of finding $\mathbb{P}^S(\tilde{A}_S(T) \geq \tilde{K})$, where $\tilde{K}$ is a modified strike $\frac{e^{-rT} \cdot K}{S_{\$}(0)}$, and the average asset $\tilde{A}$ has a modified weighting $\tilde{\mu}(t) = \mu(T-t)$. For the case of the continuous exponential weighting, we have

$$\tilde{\mu}(t) = \frac{1}{T}e^{-rt} = e^{-rT}\frac{1}{T}e^{r(T-t)} = e^{-rT} \cdot \mu(-r, t), \tag{6.5}$$

and the corresponding $\tilde{\Delta}^S(t)$ is given by

$$\tilde{\Delta}^S(t) = \int_t^T \tilde{\mu}(t)\,dt = \int_t^T \frac{1}{T} \cdot e^{-rs}\,ds = \frac{1}{rT} \cdot (e^{-rt} - e^{-rT}). \tag{6.6}$$

In the geometric Brownian motion model, we can use the partial differential equations (4.17) and (4.18) with the above modified parameters to compute the probabilities $\mathbb{P}^T(A_{B^T}(T) \geq K)$ and $\mathbb{P}^A(A_{B^T}(T) \geq K)$. As seen from equation (6.5), we can also use the results obtained from the Laplace transform by changing the sign of the interest rate $r \to -r$, and using the modified strike $\frac{K}{S_{\$}(0)}$ (the factor e^{-rT} drops out). Table 6.1 gives the prices computed from the Black–Scholes representation together with the hedging positions for the fixed strike Asian options for the set of widely used test parameters.

LEMMA 6.3. *The condition*

$$\mathcal{L}^T\left(\frac{S_{B^T}(T-t)}{S_{B^T}(0)}\right) = \mathcal{L}^S\left(\frac{B_S^T(T)}{B_S^T(t)}\right) \tag{6.7}$$

implies

$$\mathcal{L}^T(A_{B^T}(T)) = \mathcal{L}^S(\tilde{A}_S(T) \cdot S_{B^T}(0)). \tag{6.8}$$

TABLE 6.1.
List of Values Corresponding to Traditionally Used Test Parameters

r	σ	T	$S_\$(0)$	$\mathbb{P}^A$	$\mathbb{P}^T$	$\bar{\Delta}^S$ $\frac{1}{rT}(1-e^{-rT})$	Hedge $\bar{\Delta}^S\mathbb{P}^A$	Price $\bar{\Delta}^S S\mathbb{P}^A - e^{-rT}K\mathbb{P}^T$
0.05	0.5	1	1.9	0.5102	0.3955	0.9754	0.4977	0.1932
0.05	0.5	1	2.0	0.5800	0.4652	0.9754	0.5657	0.2464
0.05	0.5	1	2.1	0.6447	0.5332	0.9754	0.6289	0.3062
0.02	0.1	1	2.0	0.5762	0.5535	0.9901	0.5705	0.0560
0.18	0.3	1	2.0	0.7225	0.6609	0.9152	0.6612	0.2184
0.0125	0.25	2	2.0	0.5564	0.4751	0.9876	0.5495	0.1723
0.05	0.5	2	2.0	0.6129	0.4512	0.9516	0.5833	0.3501

Notes: The strike K is set to 2. The values for probabilities $\mathbb{P}^A$ and $\mathbb{P}^T$ were obtained from the two corresponding partial differential equations and from the Laplace transform. The values computed by both methods agree to the first four decimal digits listed in the table.

Proof.

$$
\begin{aligned}
\mathcal{L}^T(A_{B^T}(T)) &= \mathcal{L}^T\left(\int_0^T S_{B^T}(t)\mu(t)\,dt\right)\\
&= \mathcal{L}^T\left(\int_0^T S_{B^T}(T-t)\mu(T-t)\,dt\right)\\
&= \mathcal{L}^T\left(\left[\int_0^T \frac{S_{B^T}(T-t)}{S_{B^T}(0)}\mu(T-t)\,dt\right]\cdot S_{B^T}(0)\right)\\
&= \mathcal{L}^S\left(\left[\int_0^T \frac{B_S^T(T)}{B_S^T(t)}\mu(T-t)\,dt\right]\cdot S_{B^T}(0)\right)\\
&= \mathcal{L}^S\left(\left[\int_0^T S_{B^T}(t)\mu(T-t)\,dt\right]\cdot B_S^T(T)\cdot S_{B^T}(0)\right)\\
&= \mathcal{L}^S(\tilde{A}_S(T)\cdot S_{B^T}(0)). \qquad \square
\end{aligned}
$$

REMARK 6.4. Assumption (6.7) is related to self dual processes. A process $X(t)$ is self dual if it satisfies

$$
\mathbb{E}_\tau\left[f\left(\frac{X(T)}{X(\tau)}\right)\right] = \mathbb{E}_\tau\left[\frac{X(T)}{X(\tau)}\cdot f\left(\frac{X(\tau)}{X(T)}\right)\right] \tag{6.9}
$$

for any stopping time $\tau \in [0, T]$ and nonnegative Borel function f, see for instance Tehranchi (2009). If $S_{B^T}(t)$ is self dual, which is equivalent to

$$
\mathbb{E}_\tau^T\left[f\left(\frac{S_{B^T}(T)}{S_{B^T}(\tau)}\right)\right] = \mathbb{E}_\tau^S\left[f\left(\frac{B_S^T(T)}{B_S^T(\tau)}\right)\right], \tag{6.10}
$$

and has stationary increments, then Assumption (6.7) holds. Geometric Brownian motion, a class of exponential Lévy processes whose compensators satisfy $\nu^S(x) =$

$e^{-x}\nu^T(-x)$, and Ocone martingales are examples of self dual processes. This includes stochastic volatility models whose asset return is uncorrelated with the volatility process as shown in Renault and Touzi (1996). In the context of the equivalence of the Asian options with the fixed and the floating strike, it was noted in Hoogland and Neumann (2000) and in Henderson and Wojakowski (2002) that it holds in the geometric Brownian motion model. Eberlein and Papapantoleon (2005) showed that the equivalence holds for the class of exponential Lévy processes whose compensators satisfy $\nu^S(x) = e^{-x}\nu^T(-x)$.

7. EXTENSIONS TO OTHER PRICE EVOLUTIONS

This section shows how to extend these methods to the stochastic volatility models and to the pure jump model. Both models are in general incomplete and thus we give only the price representation. Let us start with the stochastic volatility model and assume that the price process S_{B^T} follows

$$dS_{B^T}(t) = g(t, \xi(t))S_{B^T}(t)\,dW^T(t), \tag{7.1}$$

where $\xi(t)$ is a stochastic process in the form

$$d\xi(t) = \alpha(t, \xi(t))\,dt + \beta(t, \xi(t))\,dW^\xi(t). \tag{7.2}$$

We assume that the two Brownian motions $W^T(t)$ and $W^\xi(t)$ are correlated:

$$dW^T(t) \cdot dW^\xi(t) = \rho dt. \tag{7.3}$$

Since $W^T(t)$ is perfectly negatively correlated with both $W^S(t)$ and $W^A(t)$, we also have

$$dW^S(t) \cdot dW^\xi(t) = -\rho dt, \qquad dW^A(t) \cdot dW^\xi(t) = -\rho dt.$$

The prices u^T, u^S, and u^A taken under different numeraires now depend also on $\xi(t)$. The following result generalizes Theorem 4.1 to the stochastic volatility model.

THEOREM 7.1. *The function* $u^T(t, x, y, \xi) = \mathbb{E}^T[V_{B^T}(T) \mid S_{B^T}(t) = x,\ A_{B^T}(t) = y, \xi(t) = \xi]$ *satisfies partial differential equation*

$$\begin{aligned} u_t^T + \alpha u_\xi^T + \frac{1}{2}g^2x^2\left[u_{xx}^T + 2\bar{\Delta}^S u_{xy}^T + [\bar{\Delta}^S]^2 u_{yy}^T\right] + \frac{1}{2}\beta^2 u_{\xi\xi}^T + \rho\beta g x u_{x\xi}^T \\ + \rho\beta\bar{\Delta}^S g x u_{y\xi}^T = 0, \end{aligned} \tag{7.4}$$

with the terminal condition

$$u^T(T, x, y, \xi) = f^T(x, y), \tag{7.5}$$

the function $u^S(t, x, y, \xi) = \mathbb{E}^S[V_S(T) \mid A_S(t) = x,\ B_S^T(t) = y, \xi(t) = \xi]$ *satisfies partial differential equation*

$$\begin{aligned} u_t^S + \alpha u_\xi^S + \frac{1}{2}g^2\left[(x - \bar{\Delta}^S)^2 u_{xx}^S + 2y(x - \bar{\Delta}^S)u_{xy}^S + y^2 u_{yy}^S\right] + \frac{1}{2}\beta^2 u_{\xi\xi}^S \\ - \rho\beta g(x - \bar{\Delta}^S)u_{x\xi}^S - \rho\beta g y u_{y\xi}^S = 0, \end{aligned} \tag{7.6}$$

with the terminal condition

$$u^S(T, x, y, \xi) = f^S(x, y), \tag{7.7}$$

and the function $u^A(t, x, y, \xi) = \mathbb{E}^A[V_A(T) \mid S_A(t) = x, B_A^T(t) = y, \xi(t) = \xi]$ *satisfies partial differential equation*

$$\begin{aligned} u_t^A + \alpha u_\xi^A &+ \frac{1}{2}g^2x^2\big([x\bar{\Delta}^S(t) - 1]^2 \cdot u_{xx}^A + 2y[x\bar{\Delta}^S(t) - 1] \cdot u_{xy}^A + y^2[\bar{\Delta}^S(t)]^2 \cdot u_{yy}^A\big) \\ &+ \frac{1}{2}\beta^2 u_{\xi\xi}^A - \rho\beta g(\bar{\Delta}^S x - 1)u_{x\xi}^A - \rho\beta g\bar{\Delta}^S xyu_{y\xi}^A = 0, \end{aligned} \tag{7.8}$$

with the terminal condition

$$u^A(T, x, y, \xi) = f^A(x, y). \tag{7.9}$$

Proof. Heuristically it follows from the Feynman-Kac Theorem. As pointed out in the proof of Theorem 4.1, the notion of viscosity solutions needs to be employed to ensure the expectation (value function) to be a solution of the associated PDE. □

When the contract depends on the assets A and S only, we get the following reduced partial differential equations.

REMARK 7.2. The function $u^S(t, x, \xi) = \mathbb{E}^S[V_S(T) \mid A_S(t) = x, \xi(t) = \xi]$ satisfies partial differential equation

$$u_t^S + \alpha u_\xi^S + \frac{1}{2}g^2(x - \bar{\Delta}^S)^2u_{xx}^S + \frac{1}{2}\beta^2u_{\xi\xi}^S - \rho\beta g(x - \bar{\Delta}^S)u_{x\xi}^S = 0, \tag{7.10}$$

with the terminal condition

$$u^S(T, x, \xi) = f^S(x), \tag{7.11}$$

and the function $u^A(t, x, \xi) = \mathbb{E}^A[V_A(T) \mid S_A(t) = x, \xi(t) = \xi]$ satisfies partial differential equation

$$u_t^A + \alpha u_\xi^A + \frac{1}{2}g^2x^2[x\bar{\Delta}^S(t) - 1]^2u_{xx}^A + \frac{1}{2}\beta^2u_{\xi\xi}^A - \rho\beta g(\bar{\Delta}^S x - 1)u_{x\xi}^A = 0, \tag{7.12}$$

with the terminal condition

$$u^A(T, x, \xi) = f^A(x). \tag{7.13}$$

Equation (7.10) was previously studied by Fouque and Han (2003). Note that if we consider the Asian forward $F(t) = A(t) - K \cdot B^T(t)$, it has the same dynamics as the average asset A:

$$\begin{aligned} dF_S(t) &= d(A_S(t) - B_S^T(t)) \\ &= (A_S(t) - \bar{\Delta}^S(t)) \cdot g(t, \xi(t))W^S(t) - K \cdot g(t, \xi(t))B_S^T(t)W^S(t) \\ &= ((A_S(t) - K \cdot B_S^T(t)) - \bar{\Delta}^S(t)) \cdot g(t, \xi(t))W^S(t) \\ &= (F_S(t) - \bar{\Delta}^S(t)) \cdot g(t, \xi(t))W^S(t). \end{aligned} \tag{7.14}$$

Therefore the function $v^S(t, x, \xi) = \mathbb{E}^S[V_S(T) \mid F_S(t) = x, \xi(t) = \xi]$ satisfies the same partial differential equation as in (7.10) and thus it can also be used for pricing the fixed strike Asian option. However, the Asian forward F cannot be used as a numeraire as it is not strictly positive, and thus equation (7.12) cannot be used.

Let us consider Asian option pricing in a pure jump model. Let us assume that the jump process has independent and stationary increments, and let $\mu(dx, dt)$ denote a random measure associated with jumps of the price process. The random measure has the following interpretation. The quantity $\int_0^t \int_A \mu(dx, ds)$ represents the number of jumps of sizes in the set A that happened in the time interval $[0, t]$. Let us denote

$$\nu(t, A) = \mathbb{E} \int_0^t \int_A \mu(dx, ds), \tag{7.15}$$

which is the expected number of jumps of sizes in the set A. This is known as a compensator. The process

$$\int_0^t \int_{-\infty}^{\infty} (\mu(dx, dt) - \nu(dx, dt)) \tag{7.16}$$

is a martingale, and it can serve as a model of the market noise. When the process has independent and time homogeneous increments, we can also write $\nu(dx, dt) = \nu(dx)dt$, so the dynamics of the jump process are determined by the sizes of the jumps, not by time. This corresponds to a Lévy process.

Let us assume that the price process is driven by the jump process with a random measure $\mu(dx, dt)$ and a compensator $\nu(dx)$. Then both price processes $S_{B^T}(t)$ and $B_S^T(t)$ are driven by the same jumps, but if one price jumps up, the inverse price jumps down accordingly. Thus the jump measure $\mu(dx, dt)$ driving the price processes differs only in the sign of the jump: $\mu^T(dx, dt) = \mu^S(-dx, dt)$. The main difference is that the reference assets B^T and S place different intensity on the jumps, so ν^T represents the compensator associated with the reference asset B^T, while ν^S represents the compensator associated with the reference asset S. Let us find the relationship between ν^T and ν^S.

The price process $S_{B^T}(t)$ can be written as

$$dS_{B^T}(t) = \int_{-\infty}^{\infty} (e^x - 1) \cdot S_{B^T}(t-)(\mu^T(dx, dt) - \nu^T(dx)\, dt). \tag{7.17}$$

The integral is over different jump sizes. From Itô's formula, the inverse price process $B_S^T(t)$ satisfies

$$dB_S^T(t) = \int_{-\infty}^{\infty} (e^{-x} - 1) \cdot B_S^T(t-)(\mu^T(dx, dt) - e^x \nu^T(dx)\, dt), \tag{7.18}$$

or

$$dB_S^T(t) = \int_{-\infty}^{\infty} (e^x - 1) \cdot B_S^T(t-)(\mu^S(dx, dt) - e^{-x} \nu^T(-dx)\, dt) \tag{7.19}$$

after switching the sign. From the symmetry between S and B^T, this equation can be written as

$$dB_S^T(t) = \int_{-\infty}^{\infty} (e^x - 1) \cdot B_S^T(t-)(\mu^S(dx, dt) - \nu^S(dx)\, dt). \tag{7.20}$$

Comparing the two above equations, we conclude that the relationship of the Lévy measures ν^T and ν^S is given by

$$\nu^S(x) = e^{-x}\nu^T(-x). \tag{7.21}$$

$$dA_{B^T}(t) = \int_{-\infty}^{\infty} \bar{\Delta}^S(t-) \cdot (e^x - 1) \cdot S_{B^T}(t-)(\mu^T(dx, dt) - \nu^T(dx)\, dt), \tag{7.22}$$

$$dA_S(t) = \int_{-\infty}^{\infty} [A_S(t-) - \bar{\Delta}^S(t-)] \cdot (e^x - 1) \cdot (\mu^S(dx, dt) - \nu^S(dx)\, dt). \tag{7.23}$$

It is not difficult to observe that the price processes $S_A(t)$ and $B_A^T(t)$ do not have stationary increments, and thus one cannot apply the methodology associated with the Lévy processes. In this case we have to limit our analysis to the choice of B^T and S as numeraires. The analogous result to Theorem 4.1 in the pure jump model is given by the following theorem.

THEOREM 7.3. *The function $u^T(t, x, y) = \mathbb{E}^T[V_{B^T}(T) \mid S_{B^T}(t) = x, A_{B^T}(t) = y]$ satisfies partial integro-differential equation*

$$\begin{aligned} u_t^T(t, x, y) + \int_{-\infty}^{\infty} \nu^T(dz)\Big[u^T\left(t, e^{\xi}x, \bar{\Delta}^S(t)(e^z - 1)x + y\right) \\ -u^T(t, x, y) - u_x^T(t, x, y) \cdot (e^z - 1)x - u_y^T(t, x, y) \cdot \bar{\Delta}^S(t)(e^{\xi} - 1)x\Big] = 0, \end{aligned} \tag{7.24}$$

with the terminal condition

$$u^T(T, x, y) = f^T(x, y), \tag{7.25}$$

and the function $u^S(t, x, y) = \mathbb{E}^S[V_S(T) \mid A_S(t) = x, B_S^T(t) = y]$ satisfies partial differential equation

$$\begin{aligned} u_t^S(t, x, y) + \int_{-\infty}^{\infty} \nu^S(dz)\Big[u^S\left(t, e^z x - (e^z - 1)\bar{\Delta}^S(t), e^z y\right) \\ - u^S(t, x, y) - u_y^S(t, x, y) \cdot (e^z - 1)y - u_x^S(t, x, y) \cdot (x - \bar{\Delta}^S(t))(e^z - 1)\Big] = 0, \end{aligned} \tag{7.26}$$

with the terminal condition

$$u^S(T, x, y) = f^S(x, y). \tag{7.27}$$

When the contract depends on the assets A and S only, we get the following reduced equation.

REMARK 7.4. The function $u^S(t,x,y) = \mathbb{E}^S[V_S(T) \mid A_S(t) = x]$ satisfies partial integro-differential equation

$$u_t^S(t,x) + \int_{-\infty}^{\infty} \nu^S(dz)\Big[u^S\left(t, e^z x - (e^z - 1)\bar{\Delta}^S(t)\right) - u^S(t,x) - u_x^S(t,x)\cdot(x - \bar{\Delta}^S(t))(e^z - 1)\Big] = 0 \tag{7.28}$$

with the terminal condition

$$u^S(T,x) = f^S(x). \tag{7.29}$$

As the Asian forward F has the same dynamics as the average asset A, equation (7.28) can also be used for pricing the fixed strike Asian option. A similar partial integro-differential equation appeared in Vecer and Xu (2004) and more recently the efficient method how to determine the solution numerically appeared in Bayraktar and Xing (2011).

APPENDIX

Proof of Theorem 5.1. The function $u^S(t,x) = \mathbb{P}^S(A_S(T) \geq K \mid A_S(t) = x)$ satisfies the following partial differential equation

$$u_t^S(t,x) + \frac{1}{2}\sigma^2(x - \bar{\Delta}^S(t))^2 u_{xx}^S(t,x) = 0, \tag{A.1}$$

with the terminal condition

$$u^S(T,x) = f^S(x) = \mathbb{I}(x \geq K), \tag{A.2}$$

Consider the following transformation of the original equation

$$v(s,z) = u^S(s(t), z(t,x)) = u^S\left(\sigma^2(T-t), \frac{2e^{-r(T-t)}}{\sigma^2 T(x - \bar{\Delta}^S(t))}\right), \tag{A.3}$$

with

$$s(t) = \sigma^2(T-t),$$

and

$$z(t,x) = \frac{2e^{-r(T-t)}}{\sigma^2 T(x - \bar{\Delta}^S(t))}.$$

Then $v(s,z)$ satisfies the partial differential equation

$$-v_s(s,z) + \left[(\kappa+1)z - \frac{1}{2}z^2\right] v_z(s,z) + \frac{1}{2}z^2 v_{zz}(s,z) = 0 \tag{A.4}$$

with $\kappa = \frac{r}{\sigma^2}$. The boundary condition when $t = T$, $s = 0$, $z = \frac{2}{\sigma^2 T x}$ is given by

$$v(0,z) = f^S\left(\frac{2}{\sigma^2 T z}\right).$$

Define the Laplace transform of v as

$$L^S(z,\lambda) = \int_0^\infty e^{-\lambda s} v(s,z)\,ds. \tag{A.5}$$

Then

(A.6)
$$-\lambda L^S(z,\lambda) + \left[(\kappa+1)z - \frac{1}{2}z^2\right] L_z^S(s,\lambda) + \frac{1}{2}z^2 L_{zz}^S(z,\lambda) = -v(0,z) = -f^S\left(\frac{2}{\sigma^2 Tz}\right).$$

By introducing a function w by

$$L^S(z,\lambda) = e^{\frac{1}{2}z} z^{-(\kappa+1)} w(z,\lambda), \tag{A.7}$$

we obtain the Whittaker's partial differential equation

$$w_{zz}(z,\lambda) + \left[-\frac{1}{4} + \frac{\kappa+1}{z} + \frac{\frac{1}{4}-\mu^2}{z^2}\right] w(z,\lambda) = -2e^{-\frac{1}{2}z} z^{\kappa-1} f^S\left(\frac{2}{\sigma^2 Tz}\right), \tag{A.8}$$

with

$$\mu = \sqrt{\left(\kappa+\frac{1}{2}\right)^2 + 2\lambda}. \tag{A.9}$$

The Green's function for the Whittaker's equation is given by

$$G(x,y) = \frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)} \cdot M_{\kappa+1,\mu}(x\wedge y)\cdot W_{\kappa+1,\mu}(x\vee y), \tag{A.10}$$

where M and W are Whittaker functions. We can write the solution for w as

$$\begin{aligned} w(z,\lambda) &= \int_0^\infty G(x,z)\cdot\left[2e^{-\frac{1}{2}x}x^{\kappa-1} f^S\left(\frac{2}{\sigma^2 Tx}\right)\right]dx, \\ &= 2\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)} \int_0^a M_{\kappa+1,\mu}(x\wedge z)\cdot W_{\kappa+1,\mu}(x\vee z)\cdot e^{-\frac{1}{2}x}x^{\kappa-1}\,dx, \end{aligned}$$

where $a = \frac{2}{\sigma^2 KT}$. When $z \geq a$, or equivalently when $K \cdot S_\$(t) \geq \int_0^t S_\$(s)\eta(s)\,ds$, we get

$$\begin{aligned}
w(z,\lambda) &= 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \cdot W_{\kappa+1,\mu}(z) \int_0^a M_{\kappa+1,\mu}(x) \cdot e^{-\frac{1}{2}x} x^{\kappa-1}\,dx \\
&= 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \cdot W_{\kappa+1,\mu}(z) \cdot \left[\frac{e^{-\frac{1}{2}x}x^\kappa}{\Gamma\left(\frac{3}{2}+\kappa+\mu\right)} \Gamma\left(\frac{1}{2}+\kappa+\mu\right) M_{\kappa,\mu}(x)\right]_{x=0}^{x=a} \\
&= 2e^{-\frac{1}{2}a} a^\kappa \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right) \cdot M_{\kappa,\mu}(a) \cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2}+\kappa+\mu\right) \cdot \Gamma(1+2\mu)}.
\end{aligned}$$

When $z < a$, we get

$$\begin{aligned}
w(z,\lambda) &= 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \int_0^a M_{\kappa+1,\mu}(x \wedge z) \cdot W_{\kappa+1,\mu}(x \vee z) \cdot e^{-\frac{1}{2}x} x^{\kappa-1}\,dx \\
&= 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \cdot \left[\int_0^z M_{\kappa+1,\mu}(x) \cdot W_{\kappa+1,\mu}(z) \cdot e^{-\frac{1}{2}x} x^{\kappa-1}\,dx\right. \\
&\quad \left. + \int_z^a M_{\kappa+1,\mu}(z) \cdot W_{\kappa+1,\mu}(x) \cdot e^{-\frac{1}{2}x} x^{\kappa-1}\,dx\right] \\
&= 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \cdot W_{\kappa+1,\mu}(z) \cdot \left[\frac{e^{-\frac{1}{2}x}x^\kappa}{\Gamma\left(\frac{3}{2}+\kappa+\mu\right)} \Gamma\left(\frac{1}{2}+\kappa+\mu\right) M_{\kappa,\mu}(x)\right]_{x=0}^{x=z} \\
&\quad + 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \cdot M_{\kappa+1,\mu}(z) \cdot \left[-e^{-\frac{1}{2}x} x^\kappa W_{\kappa,\mu}(x)\right]_{x=z}^{x=a} \\
&= 2 \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1+2\mu)} \cdot \left[\frac{e^{-\frac{1}{2}z} z^\kappa M_{\kappa,\mu}(z) \cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2}+\kappa+\mu\right)}\right. \\
&\quad \left. + e^{-\frac{1}{2}z} z^\kappa M_{\kappa+1,\mu}(z) \cdot W_{\kappa,\mu}(z) - e^{-\frac{1}{2}a} a^\kappa M_{\kappa+1,\mu}(z) \cdot W_{\kappa,\mu}(a)\right].
\end{aligned}$$

The expression for L^S follows from $L^S(z,\lambda) = e^{\frac{z}{2}} z^{-(\kappa+1)} w(z,\lambda)$.

The case $t = 0$ implies $s = \sigma^2 T$, and $z \to \infty$. But

$$\lim_{z\to\infty} e^{\frac{z}{2}} z^{-(\kappa+1)} W_{\kappa+1,\mu}(z) = 1,$$

and thus the Laplace transform L^S simplifies to

$$L^S(\lambda) = 2e^{-\frac{a}{2}} a^\kappa \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right) \cdot M_{\kappa,\mu}(a)}{\left(\frac{1}{2} + \kappa + \mu\right) \cdot \Gamma(1 + 2\mu)}. \tag{A.11}$$

Let us determine $\mathbb{P}_t^A(A_S(T) \geq K)$. This probability can be computed by considering a contract that pays off

$$\mathbb{I}(A_S(T) \geq K) \cdot A(T) = [\mathbb{I}(A_S(T) \geq K) \cdot A_S(T)] \cdot S(T).$$

We have $u^A(t, S_A(t)) = \mathbb{P}(A_S(T) \geq K \mid A_S(t)) = u^S(t, A_S(t)) \cdot S_A(t)$. Thus we can still compute the Laplace transform using the function u^S, this time with the payoff function $f^S(x) = x \cdot \mathbb{I}(x \geq K)$. Repeating the arguments from the previous part, we get

$$\begin{aligned}
\frac{w(z, \lambda)}{S_A(t)} &= \int_0^\infty G(x, z) \cdot \left[2e^{-\frac{1}{2}x} x^{\kappa-1} f^S\left(\frac{2}{\sigma^2 T x}\right)\right] dx, \\
&= 2aK \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1 + 2\mu)} \int_0^a M_{\kappa+1,\mu}(x \wedge z) \cdot W_{\kappa+1,\mu}(x \vee z) \cdot e^{-\frac{1}{2}x} x^{\kappa-2}\, dx.
\end{aligned}$$

When $z \geq a$, we get

$$\begin{aligned}
\frac{w(z, \lambda)}{S_A(t)} &= 2aK \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1 + 2\mu)} \cdot W_{\kappa+1,\mu}(z) \int_0^a M_{\kappa+1,\mu}(x) \cdot e^{-\frac{1}{2}x} x^{\kappa-2}\, dx \\
&= 2aK \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right)}{\Gamma(1 + 2\mu)} \cdot W_{\kappa+1,\mu}(z) \cdot \left[\frac{e^{-\frac{1}{2}x} x^{\kappa-1}}{\Gamma\left(\frac{3}{2} + \kappa + \mu\right)} \left(\Gamma\left(\frac{1}{2} + \kappa + \mu\right) M_{\kappa,\mu}(x)\right.\right. \\
&\qquad \left.\left. + \, \Gamma\left(-\frac{1}{2} + \kappa + \mu\right) M_{\kappa-1,\mu}(x)\right)\right]_{x=0}^{x=a} \\
&= 2Ke^{-\frac{a}{2}} a^\kappa \cdot \frac{\Gamma\left(-\frac{1}{2} - \kappa + \mu\right) \cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2} + \kappa + \mu\right) \cdot \Gamma(1 + 2\mu)} \cdot \left[M_{\kappa,\mu}(a) + \frac{M_{\kappa-1,\mu}(a)}{-\frac{1}{2} + \kappa + \mu}\right].
\end{aligned}$$

When $z < a$, we get

$$\begin{aligned}
\frac{w(z,\lambda)}{S_A(t)} &= 2aK\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)}\int_0^a M_{\kappa+1,\mu}(x\wedge z)\cdot W_{\kappa+1,\mu}(x\vee z)\cdot e^{-\frac{1}{2}x}x^{\kappa-2}\,dx\\
&= 2aK\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)}\cdot\left[\int_0^z M_{\kappa+1,\mu}(x)\cdot W_{\kappa+1,\mu}(z)\cdot e^{-\frac{1}{2}x}x^{\kappa-2}\,dx\right.\\
&\quad\left.+\int_z^a M_{\kappa+1,\mu}(z)\cdot W_{\kappa+1,\mu}(x)\cdot e^{-\frac{1}{2}x}x^{\kappa-2}\,dx\right]
\end{aligned}$$

$$\begin{aligned}
&= 2aK\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)}\cdot W_{\kappa+1,\mu}(z)\cdot\left[\frac{e^{-\frac{1}{2}x}x^{\kappa-1}}{\Gamma\left(\frac{3}{2}+\kappa+\mu\right)}\left(\Gamma\left(\frac{1}{2}+\kappa+\mu\right)M_{\kappa,\mu}(x)\right.\right.\\
&\quad\left.\left.+\ \Gamma(-\frac{1}{2}+\kappa+\mu)M_{\kappa-1,\mu}(x)\right)\right]_{x=0}^{x=z}\\
&\quad+2aK\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)}\cdot M_{\kappa+1,\mu}(z)\cdot\left[e^{-\frac{1}{2}x}x^{\kappa-1}\left(W_{\kappa-1,\mu}(x)-W_{\kappa,\mu}(x)\right)\right]_{x=z}^{x=a}\\
&= 2aK\cdot\frac{e^{-\frac{1}{2}z}z^{\kappa-1}\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)\cdot W_{\kappa+1,\mu}(z)}{\left(\frac{1}{2}+\kappa+\mu\right)\Gamma(1+2\mu)}\cdot\left[M_{\kappa,\mu}(z)+\frac{M_{\kappa-1,\mu}(z)}{-\frac{1}{2}+\kappa+\mu}\right]\\
&\quad+2aK\cdot\frac{\Gamma\left(-\frac{1}{2}-\kappa+\mu\right)}{\Gamma(1+2\mu)}\cdot M_{\kappa+1,\mu}(z)\cdot\left[e^{-\frac{1}{2}a}a^{\kappa-1}\left(W_{\kappa-1,\mu}(a)-W_{\kappa,\mu}(a)\right)\right.\\
&\quad\left.-e^{-\frac{1}{2}z}z^{\kappa-1}\left(W_{\kappa-1,\mu}(z)-W_{\kappa,\mu}(z)\right)\right].
\end{aligned}$$

The formula for L^A follows from $L^A(z,\lambda) = e^{\frac{z}{2}}z^{-(\kappa+1)}w(z,\lambda)$. □

REFERENCES

ANDREASEN, J. (1998): The Pricing of Discretely Sampled Asian and Lookback Options: A Change of Numeraire Approach, *J. Comput. Finance* 2(1), 5–30.

BAYRAKTAR, E., and H. XING (2011): Pricing Asian Options for Jump Diffusions, *Math. Finance* 21(1), 117–143.

BOYLE, P. P., and D. EMANUEL (1980): Discretely Adjusted Option Hedges, *J. Finan. Econ.* 8(3), 259–282.

EBERLEIN, E., and A. PAPAPANTOLEON (2005): Equivalence of Fixed and Floating Strike Asian and Lookback Options, *Stoch. Process. Appl.* 115, 31–40.

EL KAROUI, N., M. JEANBLANC-PICQUÉ, and S. E. SHREVE (1998): Robustness of the Black and Scholes Formula, *Math. Finance* 8(2), 93–126.

FOUQUE, J. P., and C. H. HAN (2003): Pricing Asian Options with Stochastic Volatility, *Quant. Finance* 3, 353–362.

FRIEDMAN, A. (1975): *Stochastic Differential Equations and Applications*, Vol. 1, New York: Harcourt Brace Jovanovich Publishers.

GEMAN, H., and M. YOR (1993): Bessel Processes, Asian Option, and Perpetuities, *Math. Finance* 3, 349–375.

HENDERSON, V., and R. WOJAKOWSKI (2002): On the Equivalence of Floating and Fixed Strike Asian Option, *J. Appl. Probab.* 39, 391–394.

HOOGLAND, J. K., and C. D. D. NEUMANN (2000): Asians and Cash Dividends: Exploiting Symmetries in Pricing Theory, Working Paper, CWI, Netherlands, pp. 1–16.

INGERSOLL, J. (1987): *Theory of Financial Decision Making*, Lanham, MD: Rowman and Littlefield Publishers.

LINETSKY, V. (2004): Spectral Expansions for Asian (Average Price) Options, *Operat. Res.* 52, 856–867.

RENAULT, E., and N. TOUZI (1996): Option Hedging and Implied Volatilities in a Stochastic Volatility Model, *Math. Finance* 6(3), 279–302.

ROGERS, L., and Z. SHI (1995): The Value of an Asian Option, *J. Appl. Probab.* 32, 1077–1088.

TEHRANCHI, M. R. (2009): Symmetric Martingales and Symmetric Smiles, *Stoch. Process. Appl.* 119, 3785–3797.

VECER, J. (2001): A New PDE Approach for Pricing Arithmetic Average Asian Options, *J. Comput. Finance* 4(4), 105–113.

VECER, J. (2002): Unified Asian Pricing, *Risk* 15(6), 113–116.

VECER, J., and M. XU (2004): Pricing Asian Options in a Semimartingale Model, *Quant. Finance* 4(2), 170–175.