Abstract

This paper addresses the inventory penalty pricing problem under risk-neutral valuation principle. In particular, we consider an inventory driven by a constant replenishment rate and compound renewal demand stream (iid demand inter-arrival time and iid demand quantity), and subject to overage penalties and underage penalties. Our solution approach treats the inventory penalties of overage and underage as a series of American options and constructs auxiliary martingale processes in term of inventory process. We provide a necessary and sufficient martingale condition for general compound renewal demands. For the case with compound Poisson demands, explicit expressions of penalty functions for the underage and overage are obtained.

Keywords and Phrases: Inventory penalty, martingale, overage, underage, risk-neutral pricing.
1. INTRODUCTION

Consider an inventory with constant replenishment rate and compound renewal demand stream, and subject to an overage penalty and/or an underage penalty, both assessed against the firm. An overage penalty is assessed whenever the inventory level up-crosses a prescribed target level (or capacity level), while an underage penalty is assessed whenever a demand arrival encounters a stockout where any portion of the demand cannot be satisfied from on-hand inventory. We define the perpetual inventory penalty variable as the discounted total penalty over an infinite time horizon.

The standard valuation problem in such inventory system is the computation of the expected perpetual inventory penalty. The problem we address in this paper is the inverse valuation problem of devising proper dynamic penalty functions for assessing overage and/or underage penalties, such that the expected perpetual inventory penalty equals a prescribed value (interpreted as the present value of the penalty budget given to an inventory manager). We mention that underage penalties present a special problem. Whereas overage impacts the inventory owner alone, underage impacts both the firm (e.g., lost sales, lost goodwill) and the customer (e.g., reduced service levels with potential business losses incurred by customers). While it is reasonable to assess overage penalties in terms of inventory holding costs, it is much more challenging to assess the firm’s loss of goodwill and the costs of long-term and short-term business loss sustained by the customer. Similarly, it is also hard to valuate the full extent of overage penalty because the supplier may sustain additional losses stemming from the enforcement of supply contracts (e.g., returned items). The latter are typically ignored in supply chain studies [Zipkin (2000)].

Most inventory optimization models capture overage and underage penalties as important cost components, but they rarely address the practical problem of valuating them using real-life metrics. The traditional approach is to assume a given fixed stockout cost that is incurred as a tangible loss by each stockout [Ernst and Powell (1995)]. However, intangible losses such as goodwill are not explicitly captured [Walter and Grabner (1975)]. To the best of our knowledge, the literature on the valuation inventory penalties is relatively sparse. [Schwartz (1966), (1970)] pointed out that the effect of loss of goodwill is characterized by the fact that a disappointed customer reacts in the future to change his purchasing habits. Thus the nature of the effect is that subsequent demand is perturbed, a phenomenon quite different from having an immediate penalty cost imposed, which was defined as perturbed demand. [Schwartz (1966), (1970)] analyzed some properties of perturbed demand. In addition, some numerical studies have been carried out recently to evaluate inventory penalties. For example, in order to estimate the value of lost-sales opportunities, [Ofenbakh (2008)] analyzes historical sales data to assess the underlying lost demand. More recently, there have been some studies that meld finance models and inventory management, in which inventory is treated as a financial asset and an inventory policy as a contingent claim, and well-known financial models are applied to inventory management problems. For example, [Stowe and Su (1997)] views inventory as an option on future sales and shows how some basic inventory problems can be solved using pricing models.

Our solution approach to the inverse problem is motivated by the risk-neutral valuation framework for contingent claims, and specifically by the special case of American options. The principle of risk-neutral valuation states that the option is risk-neutral if its expected value, discounted to time 0, equals its present value at time 0. It is known that if a contingent claim is risk neutral, then the stochastic process of its discounted values over time forms an exponential martingale process over an adjusted probability space [Hull (2002)].
In this paper, we shall consider an inventory penalty under the risk-neutral principle, and treat the inventory penalties of overage and underage as a series of American options (the stopping times at which penalties are incurred are analogous to American option exercise times). To this end, we shall construct auxiliary exponential martingale processes, similarly to financial valuation models, and derive a functional equation for the expected discounted costs. The auxiliary martingale process will then be used to derive an equation in terms of the Laplace transform of the pdf function of stopping times at which the penalties are incurred and an appropriate penalty function. Finally, we will use that equation with a given present value of the total penalties over an infinite time horizon.

The main contributions of this paper are listed below:
(1) For general demand process, we create an exponential martingale process and provide a necessary and sufficient condition for its martingale property.
(2) For underage penalty, we derive an explicit expression and show that the penalty value is positively and exponentially determined by the initial inventory level and the underage size as well as at a constant rate.
(3) For overage penalty, we derive an explicit expression and show that the overage penalty value is negatively and exponentially determined by the initial inventory level, but positively and exponentially determined by the capacity level at a constant rate.

Throughout the rest of this paper, we use the following notational conventions and terminology. Let \( \mathbb{R} \) denote the set of real numbers. For any real number \( x \in \mathbb{R}, \ x^+ = \max\{x, 0\} \). The indicator function \( 1_A \) is 1 if \( A \) is true, 0 otherwise. For a random variable \( X \), its pdf (probability density function) is denoted by \( f_X(x) \), cdf (cumulative distribution function) by \( F_X(x) \) and the complementary cdf by \( \overline{F}_x(x) = 1 - F_X(x) \). For any a stochastic process \( \{X(t) : t \geq 0\} \) over a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), let \( \mathcal{F}_X(s) \) denote the \( \sigma \)-algebra generated by the random variables \( \{X(t) : t \leq s\} \). The Laplace transform of a real function \( f(x) \) is defined by
\[
\hat{f}(z) = \int_0^\infty e^{-zx} f(x) \, dx.
\]
If real functions \( f(x) \) and \( g(x) \) are defined on \([0, \infty)\), then the convolution function of \( f(x) \) and \( g(x) \) is given by
\[
(f * g)(u) = \int_0^u f(u - x) g(x) \, dx.
\]
Let \( f^{(n)}(s) \) denote the \( n \)-th fold convolution of itself. Also, we assume the continuous compound interest rate, \( r \geq 0 \) is a given constant. Hence, the present value of one unit cash flow at time \( t \) is \( e^{-rt} \).

The rest of this paper is organized as follows. Section 2 introduces assessment functions and relevant background of martingale methodology. Section 3 formulates the inventory model with general demand arrivals. Section 4, 5 and 6 treat the case with compound Poisson demands. Specifically, Section 5 investigates the underage penalty, while Section 6 studies the overage penalty. Finally, Section 8 concludes this paper.
2. ASSESSMENT FUNCTIONS

In this study, we shall be interested in assessing a discounted inventory cost function in a financially reasonable manner. One abstracted assessment problem of interest can be formulated as follows. Let \( \{Y(t) : t \geq 0\} \) be a real-valued stochastic process (e.g., the evolution of a stock price, inventory level, etc.), and let \( \{\tau_i^{(L)} : i \geq 1\} \) be a sequence of hitting time of level \( L \) by the process \( \{Y(t)\} \) at which some assessments, \( \{W_i^{(L)}\} \), take place as function of the level \( L \) and the state, \( Y(\tau_i^{(L)}) \), where the assessment may be random or deterministic. Examples include exercising an American option, incurring an underage or overage penalty in an inventory system, etc. Denote the pdf function of \( \tau_i^{(L)} \) conditioned on \( Y(0) \) by \( f_{\tau_i^{(L)}}(t \mid y) \), and its conditional Laplace transform by \( \tilde{f}_{\tau_i^{(L)}}(s \mid y) \). Then, the risk-neutral value of the action at time 0 conditional on the initial state is

\[
C(L, y) = E \left[ \sum_{i=0}^{\infty} e^{-r \tau_i^{(L)}} W_i^{(L)} \mid Y(0) = y \right].
\]

(2.1)

We shall make an effort to derive the expected discounted value of the assessment occurred at the first hitting time \( \tau_i^{(L)} \),

\[
C_1(L, y) = E \left[ e^{-r \tau_1^{(L)}} W_1^{(L)} \mid Y(0) = y \right],
\]

(2.2)

since the risk-neutral value of the action given by Eq. (2.1) can be obtained recursively via conditioning on the state at first hitting time.

A classical example of the generic standard valuation problem above is the expected discounted deficit at ruin time in a classical insurance model [Gerber and Shiu (1998)]. Specifically, let \( Y(t) \) be the surplus of an insurance firm at time \( t \), and \( Y(0) = y \) the initial surplus. The firm will encounter bankruptcy at the ruin time when its surplus falls below zero for the first time. The penalty at ruin \( W_1^{(L)} \) is a function of three variables in general: the surplus immediately before ruin, the deficit at ruin, and the time of ruin.

In the more special case that \( W_1^{(L)} = w_1(L) \) is deterministic, Eq. (2.3) reduces to

\[
C_1(L, y) = w_1(L) E \left[ e^{-r \tau_1^{(L)}} \mid Y(0) = y \right] = w_1(L) \tilde{f}_{\tau_1^{(L)}}(y \mid y).\]

A classical example of the generic standard valuation problem above is a perpetual American call option. Specifically, let \( K \) be the exercise price, and \( Y(t) \) the stock price at time \( t \). Without loss of generality, assume that \( Y(0) < K \), so that exercising the option at time \( t = 0 \) can be
excluded. Then, the payoff function is \( w(L) = [L - K]^+ = L - K \). We mention that when the risk-free interest rate is constant, risk-neutral valuation provides a well-defined and unambiguous valuation methodology for the American option under consideration [Hull (2002)].

Recall that the standard valuation problem under the risk-neutral principals is to price the action exercise of Eq. (2.1) with a given payoff function \( w(L) \), while the inverse valuation problem is to deduce \( w(L) \), given \( C(L) \). In inventory budget-planning context, the latter can be formulated as follows: an inventory manager is given a prescribed perpetual inventory penalty budget that he/she is prepared to allocate to underage penalties and overage penalties, as those relate to service level metrics of the inventory system. A risk-neutral inventory manager is then faced with the (inverse) problem of deriving the appropriate penalty functions to assess per underage/overage event occurrence, subject to the principle of risk-neutral valuation (that is, the constraint that the expected perpetual inventory penalty equals to the prescribed budget).

As shown in Eq. (2.1), the functional relationship between \( C(L) \) and \( w(L) \) is specified in terms of \( \int_{y_1}^{y_2} (r | y) \), and consequently, this Laplace transform needs to be derived. To this end, we shall use an auxiliary martingale sequence \( \{M(t) : t \geq 0\} \) to define the requisite sequence of penalty functions (see section 4). Recall that a martingale process \( \{M(t) : t \geq 0\} \) satisfies the regularity condition

\[
\mathbb{E}[M(t)] < \infty, \ t \geq 0
\]  

and the martingale property

\[
\mathbb{E}[M(t) | M(s), s \leq t] = M(s), \ s \leq t.
\]  

We shall make use of the Optional Stopping Theorem (also called the Optional Sampling Theorem). One version of the theorem is given below.

**Theorem 1 [Karr (1993)]**

Let \( \{M(t)\} \) be a martingale and let \( \tau \) be a stopping time of \( \{M(t)\} \), such that

(a) \( \mathbb{E}[\tau] < \infty \)

(b) \( \{M(t)\} \) is bounded or uniformly integrable.

Then, \( \mathbb{E}[M(\tau) | M(0)] = M(0) \).

Finally, we shall make use of the following definition

**Definition 1**

Let \( \{X(t), \ t \geq 0\} \) be a stochastic process and let \( \{\theta_i, i \geq 0\} \) be a sequence of hitting time determined by \( \{X(t), \ t \geq 0\} \). Let \( \{X_i, i \geq 0\} \) be an embedded sequence, where \( X_i = X(\theta_i) \). Then \( \{X(t)\} \) has an embedded martingale \( \{X_i\} \) if it satisfies for any \( 0 \leq i \leq j \),

\[
\mathbb{E}[X_j | X(t), t \leq \theta_i] = X_i.
\]  

\[\Box\]
The motivation for defining auxiliary martingale processes as a vehicle for solving for \( \tilde{f}_{\tau(L)}(r) \) stems from the martingale methodology applied in the context of financial derivatives valuation. In particular, the evolution of an underlying security price process, \( \{S(t)\} \), is often modeled as a geometric Brownian motion, that is, the process \( \{X(t)\} \), where \( X(t) = \log S(t) \), is a shifted Brownian motion, which is routinely used in the analysis of financial derivatives (e.g., American option), due to its continuity and independent-increment property. In a risk-neutral setting, the discounted-price process of financial derivatives over time is martingale of the form [Luenberger (1998)]

\[
M(t) = e^{-r t + g(X(t))},
\]

where \( g(x) \) is a real-valued function, and \( e^{g(X(t))} \) is the price of the financial derivatives. By the martingale property, we have \( \mathbb{E}[M(t) \mid M(0)] = M(0) \). If \( \{M(t)\} \) is a bounded process, then

\[
e^{g(X(0))} = \mathbb{E}[e^{-r \tau(L) + g(X(\tau(L)))]} = \mathbb{E}[e^{-r \tau^{(L)} + g(\log L)]} = \tilde{f}_{\tau^{(L)}}(r) e^{g(\log L)}. \tag{2.8}
\]

where the first equality is due to the Optional Stopping theorem applied to the stopping time \( \tau^{(L)} \) and the second equality follows from the identity \( S(\tau^{(L)}) = L \). Eq. (2.8) readily implies

\[
\tilde{f}_{\tau^{(L)}}(r) = e^{g(X(0)) - g(\log L)}, \tag{2.9}
\]

which shows the explicit connection between the Laplace transform of \( \tau^{(L)} \) and the martingale of Eq. (2.7), and will be used to solve for the former [cf Eq. 5.7 in Gerber and Shiu (1998)].

In this paper, we use the above methodology to identify the penalty function for overage and underage penalties, subject to the principle of risk-neutral valuation, by using martingale techniques similar to those in [Gerber and Shiu (1998)], in which a standard insurance model was studied. To this end, the inventory-level process \( \{I(t) : t \geq 0\} \) will play a role analogous to \( X(t) = \log S(t) \), and the hitting time \( \tau^{(L)} \) will map to hitting times of certain inventory levels, which incur underage and overage penalties. In our case, we select \( g(x) = c x \), where \( c \) is a real constant, so that the auxiliary process \( \{M(t), t \geq 0\} \), given by

\[
M(t) = e^{-r t + c I(t)},
\]

is a martingale. Similar auxiliary martingales are discussed in [Gerber and Shiu (1998)].

### 3. INVENTORY MODEL FORMULATIONS

In this paper we consider a continuous-review inventory system with the lost-sales stockout rule. Replenishment occurs at a constant (deterministic) rate \( \rho > 0 \). The demand stream is denoted
by \( \{(A_i, D_i) : i \geq 0\} \), where the arrival time processes \( \{A_i\} \) and the corresponding demand process \( \{D_i\} \) are mutually independent. In particular, the arrival times \( \{A_i : i \geq 0\} \) follow a renewal process with arrival rate \( \lambda \), and by convention \( A_0 = 0 \). Let \( \{T_i : i \geq 1\} \) be the interarrival renewal sequence of demands, where the \( T_i = A_i - A_{i-1} \) are iid with common density \( f_T(t) \) satisfying \( f_A(0) = 0 \). Let further \( \{N_A(t) : t \geq 0\} \) be the counting process of demand arrivals, where \( N_A(t) = \max\{n \geq 0 : A_n \leq t\} \) is the number of demand arrivals in the interval \((0, t]\) with density

\[
 f_{N_A(t)}(n) = \mathbb{P}\{A_n < t < A_{n+1}\} = \mathbb{P}\{A_n < t\} - \mathbb{P}\{A_{n+1} \leq t\} = f^*_T(n)(t) - f^*_T(n+1)(t)
\]

The corresponding individual demands \( \{D_i : i \geq 0\} \) are iid with common cumulative distribution function (cdf) \( F_D(x) \) and probability density function (pdf) \( f_D(x) \), and by convention \( D_0 = 0 \). Let the cumulative demand up to time \( t \) be \( G(t) = \sum_{i=1}^{N_A(t)} D_i \). Thus, for \( t \geq 0 \) and \( s \geq 0 \),

\[
 f_G(t)(s) = \sum_{n=1}^{\infty} \left[f_T^*(n)(t) - f_T^*(n+1)(t)\right] f_D^*(n)(s).
\]

We shall consider a basic production-inventory system \( \{I(t) : t \geq 0\} \), given by

\[
 I(t) = I(0) + \rho t - \sum_{i=1}^{N_A(t)} D_i.
\]

Note that a negative inventory level means that backordering is in effect.

Next, consider the auxiliary process \( M(t) = e^{-\rho t + c I(t)} \) where the values of variable \( c \) are selected so as to ensure that each \( \{M_i = M(A_i)\} \) is an embedded martingale process.

**Lemma 1**

The process \( \{M(t)\} \) has an embedded martingale \( \{M_i\} \) if and only if \( c \) satisfies

\[
 \tilde{f}_T(r - c \rho) \tilde{f}_D(c) = 1.
\]

**Proof.** See Appendix.

We next investigate the roots of Eq. (3.4). To this end, we define

\[
 L_T(c) = -\log \tilde{f}_T(r - c \rho)
\]

and
\[ L_D(e) = \log \tilde{f}_D(e), \quad (3.6) \]

and note that Eq. (3.4) is equivalent to

\[ L_T(e) = L_D(e). \quad (3.7) \]

The following result provides key properties of these functions.

**Lemma 2**

(a) \( L_T(e) \) is strictly decreasing and strictly concave in \( e \).

(b) \( L_D(e) \) is strictly decreasing and strictly convex in \( e \).

**Proof.** See Appendix. $\Box$

The following Proposition provides a key property for the roots of Eq. (3.4).

**Proposition 1**

Eq. (3.4) has two real roots, \( c_1 \leq 0 \) and \( c_2 \geq 0 \).

**Proof.**

Follows immediately from the Lemma 2 and the fact that \( L_D(0) = 0 < L_T(0) \) by Eqs. (3.5) and (3.6). $\Box$

Figure 1 depicts the functional form of \( L_T(e) \) and \( L_D(e) \), as well as the location of the roots, \( c_1 \) and \( c_2 \).

**Figure 1.** The Functional form of \( L_T(e) \) and \( L_D(e) \) and the two roots of \( L_T(e) = L_D(e) \).
4. COMPOUND POISSON DEMAND ARRIVALS

In this section and the rest of this paper, we consider the case where demand arrivals follow a Poisson process with arrival rate \( \lambda > 0 \), that is the density function of interarrival time \( T \) satisfies \( f_T(t) = \lambda e^{-\lambda t} \). Then Eq. (3.4) simplifies to

\[
 r - \rho c + \lambda [1 - \tilde{f}_D(c)] = 0. \tag{4.1}
\]

Eq. (4.1) is well known in the context of insurance models, where it is referred to as *Lundberg’s fundamental equation* [Gerber and Shiu (1998)]. Accordingly, we shall refer to Eq.(3.4) as the renewal extension of Lundberg’s fundamental equation.

The following example provides the two roots for the case with Poisson inter-arrival time and Exponential demands.

**Example 1 (Poisson arrival and Exponential demands)**

Consider the case with exponential demand size of parameter \( \beta > 0 \), that is \( f_D(x) = \beta e^{-\beta x} \).

Hence, \( \tilde{f}_D(c) = \frac{\beta}{\beta + c} \). Then Eq. (4.1) can be rewritten as

\[
 r - \rho c + \lambda [1 - \frac{\beta}{\beta + c}] = 0,
\]

or equivalently,

\[
 \rho c^2 + (\rho \beta - r - \lambda) c - \beta r = 0.
\]

Hence, two roots are obtained as

\[
 c_1 = \frac{r + \lambda - \rho \beta - \sqrt{(\rho \beta - r - \lambda)^2 - 4 \rho \beta r}}{2 \rho},
\]

\[
 c_2 = \frac{r + \lambda - \rho \beta + \sqrt{(\rho \beta - r - \lambda)^2 - 4 \rho \beta r}}{2 \rho}.
\]

With two roots \( c_1 \) and \( c_2 \), we define the two processes,

\[
 M_1(t) = e^{-r t + c_1 I(t)} \tag{4.2}
\]

and

\[
 M_2(t) = e^{-r t + c_2 I(t)}. \tag{4.3}
\]
**Proposition 2**

If the demand arrivals follow a Poisson process and \( c_1 \leq 0 \) and \( c_2 \geq 0 \) satisfy Eq.(4.1), then \( \{M_1(t)\} \) and \( \{M_2(t)\} \) are both continuous martingales in \( t \).

**Proof.**

The proof is completed by following the memoryless property of Poisson process and replacing \( A_i \) and \( A_j \) with \( t \) and \( s \), respectively, in the proof of Lemma 1.

By Proposition 2, both \( \{M_1(t)\} \) and \( \{M_2(t)\} \) are martingale processes if \( c_1 \) and \( c_2 \) satisfy Eq.(4.1), or Eq.(3.4) in more general. However, the martingale property does not hold always while a stopping time (e.g, underage or overage stopping time) is considered. Fortunately, in our cases, the martingale property is still holding if we associate \( \{M_1(t)\} \) and \( \{M_2(t)\} \) with underage and overage stopping time, respectively. In the following, we shall associate \( \{M_1(t)\} \) with underage in section 5, while associate \( \{M_2(t)\} \) with overage in section 6.

## 5. UNDERAGE PENALTY FUNCTION

In this section, we study the basic production-inventory systems under the lost sales, where unmet demand will be lost subject to an underage (or lost-sale) penalty. An underage penalty is assessed whenever the inventory level hits 0 (at hitting times). We are interested in deriving the explicit function for the underage penalty. To this end, we shall take the advantage of the aforementioned process \( \{M_1(t)\} \), by which we derive some cost properties pertaining to the system.

Under lost sale operation, all the shortage if any will be lost, which is interpreted as lost sales. The *lost-sales size* is denoted by

\[
L(A_i) = [D_i - I(A_i^-)]^+.
\]

The inventory process under lost sales, \( \{I(t) : t \geq 0\} \), is given by the following equations

\[
I(t) = u + pt - \sum_{i=1}^{N_d(t)} [D_i - L(A_i^-)],
\]

where the summation term represents the cumulative amount of demands that have been satisfied. Then, the stopping times of underage in the process defined above are denoted by

\[
\tau_{k}^{(0)} = \inf\{A_i > \tau_{k-1}^{(0)} : L(A_i) > 0\}.
\]

where \( \tau_{k}^{(0)} = 0 \) while \( k = 0 \) by convention, the lost-sale size \( L(A_k) \) is given by Eq. (5.1), and we use superscript “0” to denote the hitting inventory level for lost-sale occurrence. Figure 2 illustrates a sample path of the inventory level process.
The following theorem provides the martingale property of \( \{M_1(t)\} \).

**Theorem 2**

The process \( \{M_1(t)\} \) defined by Eq. (4.2) is a martingale process over \([0, \tau_1^{(0)}]\).

**Proof.**

By Proposition 2, we have \( \{M_1(t)\} \) is martingale. Furthermore, if \( \tau_1^{(0)} < \infty \), we have \( 0 \leq M_1 \leq 1 \) by Eq. (4.2) since \( -rA_i + c_i I(A_i) \leq 0 \). Applying the optional sampling theorem, the proof is completed.

Without loss of generality, we denote the underage penalty for the \( k \)-th underage by \( w_k^{(0)} \left( I(\tau_k^{(0)}) - L(\tau_k^{(0)}) \right) \) as a function of the inventory level right before underage and the corresponding lost-sales size. It is of our interest to derive an explicit expression for the underage penalty. To this end, we shall first study the expected discounted penalty for the first underage in subsection 5.1, and then extend to the sequential underage over infinite time horizon in subsection 5.2, and finally derive for the attendant penalty values.

### 5.1 The First Underage

In this subsection, our main effort is to derive the expected discounted cost for the first underage. To this end, we shall reconsider the basic inventory process given by Eq. (5.3). Figure 3 illustrates a sample path of the inventory process till the stopping time of first underage.
In light of Theorem 2, one immediately has
\[ M_1(0) = \mathbb{E}\left[ M_1(\tau_1^{(0)}) \mid I(0) = u \right], \]
which can be further written as
\[ e^{c_1 u} = \mathbb{E}\left[ e^{-r \tau_1^{(0)} + c_1 I(\tau_1^{(0)})} \mid I(0) = u \right]. \quad (5.4) \]

**Example 2 (Exponential demands)**
Consider the case in which the demand arrivals follow a Poisson distribution and the demand size has exponential distribution with pdf function \( f_D(x) = \beta e^{-\beta x} \), where \( \beta > 0 \). Two roots to Eq. (3.4) for this case have been obtained in Example 1. We then have the following result.

**Proposition 3**
If the demand size is Exponential with parameter \( \beta > 0 \), then for any \( u \geq 0 \), the following is true
\[ \mathbb{E}\left[ e^{-r \tau_1^{(0)}} \mid I(0) = u \right] = \frac{\beta + c_1}{\beta} e^{c_1 u}. \quad (5.5) \]

**Proof.** See Appendix. \( \Box \)

In particular, if \( u = 0 \) in Eq. (5.5), one has
\[ \mathbb{E}[e^{-r \tau_1^{(0)}} \mid I(0) = 0] = \frac{\beta + c_1}{\beta}. \quad (5.6) \]

More specifically, if the penalty function is a function of the underage size denoted by \( w^{(0)}(L(\tau_1^{(0)})) \), then the expected discounted penalty for the first lost sale can be written as

![Figure 3. A sample path of the inventory level](image-url)
\[ \mathbb{E}[e^{-r \tau_i(0)} w^{(0)}(I(\tau_1^{(0)}) \mid I(0) = u] = \int_0^\infty w^{(0)}(y) f_D(y) dy \mathbb{E}\left[ e^{-r \tau_i(0)} \mid I(0) = u \right] \]
\[ = \frac{\beta + c_1}{\beta} e^{c_1 u} \mathbb{E}[w^{(0)}(D)] \]  
(5.7)

As shown in Eq. (5.7), the expected discounted penalty of the first underage is exponentially decreasing in the initial inventory level \( u \). The increasing rate is determined uniquely by the root value \( c_1 \).

### 5.2 The Underage Sequence

In this subsection, we shall study the underage sequence over the time. For \( k > 1 \), we have
\[
\mathbb{E}\left[ e^{-r \tau_k^{(0)}} \mid I(0) = u \right] = \mathbb{E}\left[ e^{-r \tau_{k-1}^{(0)}} \mid I(0) = u \right] \mathbb{E}\left[ e^{-r [\tau_k^{(0)} - \tau_{k-1}^{(0)}]} \mid I(0) = u \right]
\]
\[ = \mathbb{E}\left[ e^{-r \tau_{k-1}^{(0)}} \mid I(0) = u \right] \mathbb{E}\left[ e^{-r \tau_1^{(0)}} \mid I(0) = 0 \right]. \]  
(5.8)

\[ = \frac{\beta + c_1}{\beta} \mathbb{E}\left[ e^{-r \tau_{k-1}^{(0)}} \mid I(0) = u \right] \]

where the last equality holds by Eq. (5.6). A combination of Eq. (5.8) and Eq. (5.5) yields
\[
\mathbb{E}\left[ e^{-r \tau_k^{(0)}} \mid I(0) = u \right] = \left( \frac{\beta + c_1}{\beta} \right)^k e^{c_1 u}. \]  
(5.9)

We then have the following result.

**Theorem 3**

If the prescribed perpetual underage penalty budget is \( p_U \), the underage penalty \( w_k^{(0)}(y) \) are identical over \( k \) such that \( w_k^{(0)}(y) = w^{(0)}(y) \) as a function of the underage size, then
\[
w^{(0)}(y) = p_U e^{-c_1 y} \left[ e^{-c_1 y} - 1 \right]. \]  
(5.10)

**Proof.**

By Eq. (5.8), one has
\[
\mathbb{E}\left[ \sum_{k=1}^\infty e^{-r \tau_k^{(0)}} w_k^{(0)}(Y) \mid I(0) = u \right] = \frac{\beta + c_1}{c_1} e^{c_1 u} \mathbb{E}[w^{(0)}(D)] \]  
(5.11)

- 12 -
By the principle of risk-neutral valuation, the right hand side of Eq. (5.11) equals to $p_U$, which is the amount the manager would pay for underage in all at time 0, then

$$
\mathbb{E} [w(0)(D)] = p_U \frac{-c_1}{\beta + c_1} e^{-c_1 u}
$$

Note that $\mathbb{E} [w(0)(D)] = \beta \bar{w}(0)(\beta)$. The equation above yields

$$
\bar{w}(0)(\beta) = p_U \frac{-c_1}{(\beta + c_1) \beta} e^{-c_1 u}
$$

(5.12)

Finally, taking inverse Laplace transform of Eq. (5.12), we complete the proof. \hfill \Box

Theorem 3 implies that the penalty value is positively exponentially determined by the initial inventory level and the underage size as well at degree of $c_1$.

6. OVERAGE PENALTY FUNCTIONS

In this section, we consider production-inventory systems under the Base-stock rule, where the inventory has an upper threshold level (or referred to as a capacity level), $S > 0$, such that an overage penalty is assessed whenever the inventory level hits $S$ (at hitting times). Replenishment is governed by the base-stock policy as follows: it proceeds at a rate $\rho > 0$ while the inventory level is below the $S$, and is suspended, otherwise. All the shortage of the demand is backordered. Then, the process of the inventory with threshold level, $\{I(t) : t \geq 0\}$, is given by the following equations

$$
I(t) = u + \int_0^t \rho 1_{\{I(z) < S\}} \, dz - \sum_{i=1}^{N_A(t)} D_i. \quad (6.1)
$$

For $u < S$, we define a sequence of stopping times when inventory hits the capacity level $S$,

$$
\tau_k^{(S)} = \inf\{ t > \tau_{k-1}^{(S)} : I(t) = S \}, \quad (6.2)
$$

where $\tau_0^{(S)} = 0$, by convenience, if $k = 0$. Note that for each $k$, since $\{I(t)\}$ is jump free upward, we have

$$
I(\tau_k^{(S)}) = S. \quad (6.3)
$$
In the following, we are interested in deriving the penalty function for overage.

Let \( \alpha_k^{(s)} = A_{N(\tau_k^{(s)})+1} \) denote the first demand arriving time after the \( k \)-th overage occurrence. Then the corresponding replenishment idle time is

\[
T_k^{(s)} = \alpha_k^{(s)} - \tau_k^{(s)}.
\] (6.4)

By the memoryless property of exponential distribution, we have that \( \{T_k^{(s)}\} \) are iid, and follow an exponential distribution with parameter \( \lambda^{-1} \). Whence, for \( k = 1, 2, 3, \ldots \), it holds that

\[
\mathbb{E} \left[ e^{-r T_k^{(s)}} \right] = \frac{\lambda}{\lambda + r}.
\] (6.5)

Let the overage penalty for the \( k \)-th overage be parameterized by the target level and denoted by \( w_k^{(s)} \). Then the expected discounted penalty for overage sequence is

\[
\mathbb{E} \left[ \sum_{k=1}^{\infty} w_k^{(s)} \cdot e^{-r \tau_k^{(s)}} | I(0) = u \right] = \sum_{k=1}^{\infty} w_k^{(s)} \mathbb{E} \left[ e^{-r \tau_k^{(s)}} | I(0) = u \right].
\] (6.6)

In the following, we first study the expected discounted penalty for the first overage, then extend to the expected discounted penalty for the overage sequence over an infinite time horizon.

6.1 The First Overage

Shown in Figure 5 is a sample path of the inventory level process until the first overage occurrence.
Considering \( \{M_2(t)\} \), we have the following theorem.

**Theorem 4**

The process \( \{M_2(t)\} \) is a martingale over \( [0, \tau_1^{(S)}] \).

**Proof.**

By Proposition 2, we have \( \{M_2(t)\} \) is martingale. Furthermore, if \( \tau_1^{(S)} < \infty \), we have

\[ 0 \leq M_2(t) \leq e^{c_2u} \]

uniformly in \( t \) by Eq. (4.3). Applying the optional sampling theorem, the proof is completed.

By Theorem 4 we immediately have \( M_2(0) = \mathbb{E}[M_2(t) \mid I(0) = u] \). Specifically, it readily follows if we set \( t = \tau_1^{(S)} \),

\[
e^{c_2u} = \mathbb{E}\left[e^{-r\tau_1^{(S)} + c_2I(\tau_1^{(S)})} \mid I(0) = u\right] = e^{c_2S} \mathbb{E}\left[e^{-r\tau_1^{(S)}} \mid I(0) = u\right].
\]

where the second equality follows from Eq. (6.3). It implies

\[
\mathbb{E}\left[e^{-r\tau_1^{(S)}} \mid I(0) = u\right] = e^{c_2[u-S]}
\]

(6.8)

Then the expected discount penalty for the first overage is

\[
\mathbb{E}\left[w^{(S)} e^{-r\tau_1^{(S)}} \mid I(0) = u\right] = w^{(S)} \mathbb{E}\left[e^{-r\tau_1^{(S)}} \mid I(0) = u\right] = w^{(S)} e^{c_2[u-S]}
\]

(6.9)

where the second equality holds by Eq. (6.8).

We shall mention that Eqs. (6.7) - (6.9) remain valid even if \( u \leq 0 \) or \( u = S \). The condition for the initial inventory level is \( u \leq S \).
6.2 The Overage Sequence

For $k > 1$, we have

$$
E\left[ e^{-r\tau_{k}^{(S)}} | I(0) = u \right] = E\left[ e^{-r\tau_{k-1}^{(S)}} | I(0) = u \right] E\left[ e^{-r\tau_{k-1}^{(S)} + \tau_{k-1}^{(S)}} | I(0) = u \right]
$$

$$
= \lambda \int_{0}^{\infty} f_{D}(x) e^{c_{2}[x-S]} dx E\left[ e^{-r\tau_{k-1}^{(S)} | I(0) = u \right]
$$

$$(6.10)$$

Combining Eq. (6.10) and Eq. (6.8) yields,

$$
E\left[ e^{-r\tau_{k}^{(S)}} | I(0) = u \right] = \gamma^{k-1} e^{c_{2}[u-S]} ,
$$

$$(6.11)$$

where

$$
\gamma = \frac{\lambda}{\lambda + r} e^{-c_{2}S} \tilde{f}_{D}(-c_{2}) .
$$

$$(6.12)$$

Note that $0 < \gamma < 1$ since $0 < \frac{\lambda}{\lambda + r} < 1$, $0 < e^{-c_{2}S} < 1$ and $0 < \tilde{f}_{D}(-c_{2}) < 1$.

Substituting Eq. (6.11) into Eq., we have

$$
E\left[ \sum_{k=1}^{\infty} w_{k}^{(S)} \cdot e^{-r\tau_{k}^{(S)}} | I(0) = u \right] = e^{-c_{2}[u-S]} \sum_{k=1}^{\infty} w_{k}^{(S)} \gamma^{k-1} .
$$

$$(6.13)$$

In particular, if the overage penalty is identical over $k$, then we have the following result.

**Theorem 5**

If the prescribed perpetual overage penalty budget is $p_{O}$, and the overage penalty is identical over $k$ and denoted by $w_{k}^{(S)} = w^{(S)}$, then the overage penalty for overage level $S$ is

$$
w^{(S)} = p_{O} (1 - \gamma) e^{c_{2}[S-u]} .
$$

$$(6.14)$$

where $\gamma = \frac{\lambda}{\lambda + r} e^{-c_{2}S} \tilde{f}_{D}(-c_{2})$.

**Proof.**

Setting $w_{k}^{(S)} = w^{(S)}$ in Eq. (6.13), one has
By the principle of risk-neutral valuation, the most right hand side of Eq. (6.15) equals to \( p_O \), which is the amount the manager would pay for overage penalties in all at time 0. Therefore

\[
\mathbb{E} \left[ \sum_{k=1}^{\infty} w_k^{(S)} \cdot e^{-r \tau_k^{(S)}} \mid I(0) = u \right] = e^{r [u-S]} w^{(S)} \sum_{k=1}^{\infty} \gamma^{k-1} = \frac{e^{r [u-S]} w^{(S)}}{1 - \gamma}.
\] (6.15)

The result readily follows by simple algebra.

Theorem 5 implies that the overage penalty value is negatively exponentially determined by the initial inventory level \( u \), but positively exponentially determined by \( S \) at degree of \( e_2 \). In particular, if \( I(0) = u = S \), then \( w^{(s)} = p_O \left( 1 - \frac{\lambda}{\lambda + r} e^{-c_2 S} \tilde{f}_D(-c_2) \right) \) which is exponentially increasing in \( S \).

7. CONCLUDING REMARKS

As an interface between finance and inventory management, a major goal of this investigation is to formulate the penalty functions and derive closed-form expressions for overage and underage penalties under risk-neutral valuation principle in a quantitative and methodological effort. In particular, we consider an inventory driven by a constant replenishment rate and compound renewal demand stream, and subject to overage penalties and underage penalties.

Our solution approach treats the inventory penalties of overage and underage as a series of American options. To price such options, we construct auxiliary exponential martingale processes in term of inventory process. The auxiliary martingale process will then be used to achieve some intermediate results that help derive the expected discounted penalty costs. Under risk-neutral valuation principle, we derive the penalty function with a prescribed perpetual inventory (overage or underage) penalty budget. In particular, for general compound renewal demands, we obtain the necessary and sufficient martingale condition. We then prove that there are two roots to this equation, one is positive while the other is negative. Accordingly, we associate the martingale constructed based on the negative root with the underage discussion, while the martingale constructed based on the positive root with the overage. For the case with compound Poisson demands, explicit expressions for the underage penalty and overage penalty are obtained.
References
APPENDIX

Proof of Lemma 1
Substituting Eq. (3.3) into Eq. (2.10) and noting that I(t) and M(t) determine each other for each t, we can rewrite the left hand side of Eq. (2.6) for any 0 ≤ i ≤ j as

\[ \mathbb{E}[M_j \mid M(t), t \leq A_i] = \mathbb{E}\left[ e^{-r A_j + c I(A_j)} \mid M(t), t \leq A_i \right] \]

\[ = e^{-r [A_j - A_i] + c [I(A_j) - I(A_i)]} \mathbb{E}[I(t), t \leq A_i] \]

\[ = M_i \mathbb{E}\left[ e^{-r [A_j - A_i] + c (\rho [A_j - A_i] - [S(A_j) - S(A_i)])} \mid I(A_i) \right] \]

\[ = M_i \mathbb{E}\left[ e^{-(r - c \rho) \sum_{k=i+1}^j T_k} \right] \mathbb{E}\left[ e^{-c \sum_{k=i+1}^j D_k} \right] \]

(8.1)

For the embedded martingale condition to hold, it is necessary and sufficient that the right-most hand side of Eq. (8.1) equal \( M_i \), which is equivalent to the condition

\[ 1 = \mathbb{E}\left[ e^{-(r - c \rho) \sum_{k=i+1}^j T_k} \right] \mathbb{E}\left[ e^{-c \sum_{k=i+1}^j D_k} \right] \]

\[ = \tilde{f}_T^{j-i} (r - c \rho) \tilde{f}_D^{j-i} (c) \]

Eq. (3.4) now readily follows, since \( \tilde{f}_T (r - c \rho) \tilde{f}_D (c) \) is non-negative.

Proof of Lemma 2
To prove part (a), we show that the derivatives of \( L_T(c) \) satisfy

\[ L_T'(c) = -\frac{\partial}{\partial c} \tilde{f}_T (r - c \rho) < 0 \]

(8.2)

\[ L_T''(c) = -\frac{\tilde{f}_T (r - c \rho) \frac{\partial^2}{\partial c^2} \tilde{f}_T (r - c \rho) - \left( \frac{\partial}{\partial c} \tilde{f}_T (r - c \rho) \right)^2}{[\tilde{f}_T (r - c \rho)]^2} < 0. \]

(8.3)

Eq. (8.2) follows from the facts
\[ \tilde{f}_T(r - c\rho) = \mathbb{E}[e^{-(r-c\rho)T_i}] > 0, \quad (8.4) \]
\[ \frac{\partial}{\partial c} \tilde{f}_T(r - c\rho) = \rho \int_0^\infty t e^{-(r-c\rho)t} f_T(t) \, dt = \rho \mathbb{E}[T_i^2 e^{-(r-c\rho)T_i}] > 0 \quad (8.5) \]

where Eq. (8.4) is an immediate consequence of the Laplace transform and Eq. (8.5) follows by the Leibniz integral rule.

To prove Eq. (8.3), note that the denominator is positive, so it remains to show that the numerator is positive too. Differentiating Eq. (8.5) with the aid of the Leibniz integral rule yields

\[ \frac{\partial^2}{\partial c^2} \tilde{f}_T(r - c\rho) = \rho^2 \int_0^\infty t^2 e^{-(r-c\rho)t} f_T(t) \, dt = \rho^2 \mathbb{E}[T_i^2 e^{-(r-c\rho)T_i}] \quad (8.6) \]

Substituting Eqs. (8.4), (8.5) and (8.6) into the numerator of Eq. (8.3), the latter becomes

\[ \mathbb{E}[e^{-(r-c\rho)T_i}] \rho^2 \mathbb{E}[T_i^2 e^{-(r-c\rho)T_i}] - \rho^2 \mathbb{E}[T_i e^{-(r-c\rho)T_i}] \quad (8.7) \]

Finally, applying the Cauchy-Schwarz inequality [Karr (1993)] to the product of expectations in the first term of Eq. (8.7) results in the inequality,

\[ \mathbb{E}[e^{-(r-c\rho)T_i}] \mathbb{E}[T_i^2 e^{-(r-c\rho)T_i}] > \mathbb{E}[T_i e^{-(r-c\rho)T_i}], \quad (8.8) \]

where the strict inequality follows from the fact that the random variables on the left-hand side above are not proportional to each other. Eq. (8.8) establishes that the numerator of Eq. (8.3) is positive, thereby completing the proof of part (a).

To prove part (b), we apply an argument similar to that for part (a). Accordingly, we show that the derivatives of \( L_D(c) \) satisfy

\[ L'_D(c) = \frac{\tilde{f}_D'(c)}{\tilde{f}_D(c)} < 0 \quad (8.9) \]
\[ L''_D(c) = \frac{\tilde{f}_D''(c)\tilde{f}_D(c) - \tilde{f}_D'(c)^2}{[\tilde{f}_D(c)]^2} > 0 \quad (8.10) \]

Eq. (8.9) follows from the facts

\[ \tilde{f}_D(c) = \mathbb{E}[e^{-cD_i}] = \int_0^\infty e^{-cD} f_D(x) \, dx > 0 \quad (8.11) \]
\[ \tilde{f}_D'(c) = \int_0^\infty (-x)e^{-cD} f_D(x) \, dx = \mathbb{E}[-D_i e^{-cD_i}] < 0. \quad (8.12) \]

where Eq. (8.11) is an immediate consequence of the Laplace transform and Eq. (8.12) follows by the Leibniz integral rule.
To prove Eq. (8.10), note that the denominator is positive, so it remains to show that the numerator is positive too. Differentiating Eq. (8.12) with the aid of the Leibniz integral rule yields,

$$
\tilde{f}''_D(c) = \int_0^\infty x^2 e^{-cx} f_D(x) \, dx = \mathbb{E}\left[ D_i^2 e^{-cD_i} \right] > 0.
$$

(8.13)

Substituting Eqs. (8.11), (8.12) and (8.13) into the numerator of Eq. (8.10), the latter becomes

$$
\mathbb{E}\left[ D_i^2 e^{-cD_i} \right] \mathbb{E}\left[ e^{-cD_i} \right] - \mathbb{E}^2[D_i e^{-cD_i}]
$$

(8.14)

Finally, applying the Cauchy-Schwarz inequality to the first term of Eq. (8.14) results in the inequality,

$$
\mathbb{E}\left[ D_i^2 e^{-cD_i} \right] \mathbb{E}\left[ e^{-cD_i} \right] > \mathbb{E}^2[D_i e^{-cD_i}]
$$

(8.15)

where the strict inequality follows from the fact that the random variables on the left-hand side above are not proportional to each other. Eq. (8.15) establishes that the numerator of Eq. (8.10) is positive, thereby completing the proof of part (b).

\[ \square \]

Proof of Proposition 3

For ease of exposition, let \( f(x, y, t \mid u) \) denote the joint density function of \( I\left(\tau^{(0)}_1 - \right) \), \( L(\tau^{(0)}_1) \) and \( \tau^{(0)}_1 \), where \( x, y, t \) are non-negative real numbers. Define

\[
 f(x, y \mid u) = \int_0^\infty e^{-r t} f(x, y, t \mid u) \, dt,
\]

and the marginal density functions are denoted respectively by

\[
 f(x \mid u) = \int_0^\infty f(x, y \mid u) \, dy
\]

and

\[
 f(y \mid u) = \int_0^\infty f(x, y \mid u) \, dx.
\]

Therefore, one has the following relationship

\[
 \mathbb{E}\left[ e^{-r \tau^{(0)}_1} \mid I(0) = u \right] = \int_0^\infty \int_0^\infty \int_0^\infty e^{-r t} f(x, y, t \mid u) \, dy \, dx \, dt
\]

(8.16)

By the definition of conditional probability, one further has

\[
 f(x, y \mid u) = f(x \mid u) \frac{f_D(x + y)}{1 - F_D(x)}.
\]

Note that, by the memorylessness property of exponential distribution the second term on the right hand side above can be simplified as

\[
 \frac{f_D(x + y)}{1 - F_D(x)} = f_D(y).
\]

Hence,
Then,
\[
\begin{align*}
\mathbb{E}\left[ e^{-r\tau_1^{(0)}+c_1 I_1^{(0)}} \mid I(0) = u \right] &= \int_0^\infty \int_0^\infty \int_0^\infty e^{-r t - c_1 y} f(x, y, t \mid u) \, dx \, dy \, dt \\
&= \int_0^\infty \int_0^\infty e^{-c_1 y} f(x, y \mid u) \, dx \, dy \\
&= \int_0^\infty e^{-c_1 y} f_D(y) \, dy \mathbb{E}\left[ e^{-r\tau_1^{(0)}} \mid I(0) = u \right] \\
&= \frac{\beta}{\beta + c_1} \mathbb{E}\left[ e^{-r\tau_1^{(0)}} \mid I(0) = u \right]
\end{align*}
\]
where the fourth equation holds by Eq. (8.16). By Eq. (5.4), the above equation equals \( e^{c_1 u} \), which therefore completes the proof. 
\[\Box\]