

考虑分红情形下QLBS模型的期权定价与对冲

黄瑜

广东工业大学数学与统计学院, 广东 广州

收稿日期: 2026年3月8日; 录用日期: 2026年4月2日; 发布日期: 2026年4月10日

摘要

本文针对标的资产分红的实际市场特征, 对基于强化学习的QLBS期权定价与对冲模型进行扩展。推导了连续分红情形下QLBS模型的离散时间理论框架, 重构了含分红的自融资复制组合、奖励函数与最优对冲解析解, 并通过蒙特卡洛模拟与B样条基函数逼近完成数值实现。数值实验结果表明, QLBS^(q)模型在不同分红率下均可稳定逼近含分红BSM模型的定价结果, 有效提升了模型对实际市场的适配性。

关键词

QLBS模型, 欧式期权定价, 连续分红, 强化学习

Option Pricing and Hedging of the QLBS Model under Dividend Scenarios

Yu Huang

School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou
Guangdong

Received: March 8, 2026; accepted: April 2, 2026; published: April 10, 2026

Abstract

This paper addresses the limitations of the the actual market characteristics of underlying asset dividends, extending the option pricing and hedging model (QLBS) based on reinforcement learning. Derived the discrete-time theoretical framework of the QLBS model under continuous dividend scenarios , and the self-financing replication portfolio with dividends, the reward function. And the optimal hedging analytical solution are reconstructed. Numerical implementation is achieved through Monte Carlo simulation and B-splines basis function approximation. Numerical experimental results show that the $QLBS^{(q)}$ model can stably approximate the pricing results of the dividend-paying BSM model under different dividend rates, effectively improving the model's adaptability to the real market.

Keywords

QLBS Model, European Option Pricing, Continuous Dividends, Reinforcement Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



1. 引言

自Black 和Scholes [1]引入经典的BSM模型以来,对期权的定价问题成为金融领域关注的研究领域之一。几十年来,学者们在BSM模型的基础上进行了大规模的研究。比如考虑到期权有效期内标的资产可能支付股息,许多学者开始探讨带有股息支付的期权定价问题。Merton [2]考虑支付连续股息收益率的标的资产,并推导出对应的B-S公式。Krausz [3], Broadie和Detemple [4]给出了适用于支付连续股息的美式和欧式期权定价公式。随着期权市场的复杂性和不确定性不断增加,当实际资产动态与模型假设不完全一致时,最终的价格和对冲可能会受到模型错误所产生的影响。强化学习 [5]作为一种数据驱动的优化方法,近年来在金融领域得到了广泛关注。这些研究在提高期权的定价和对冲的准确性方面取得了卓越的成果。

对于欧式期权, Halperin [6]提出了一个基于Q-learning的离散时间期权定价模型,称为QLBS。

该模型直接从交易数据中学习得到期权定价和对冲策略,为利用强化学习解决期权定价和对冲问题提供了一种新的思路。随后在该模型的基础上探索了Fitted Q Iteration、逆向强化学习及其在对冲组合定价中的应用 [7]。Stoiljkovic [8]的最近研究表明,QLBS模型在欧式看跌期权不同状态变量和场景下都能准确地逼近BSM理论数值。基于这些已有的研究成果,我们给出在标的资产按连续股利收益率 q 分红时,QLBS模型的完整离散时间推导,扩展研究了股票支付连续红利的期权定价。将其纳入模型可以更好地反映实际市场情况,进一步提升模型的应用价值。

本文的概述如下:第2节介绍了BSM模型以及QLBS模型给出的最优期权定价和对冲解析解,第3节利用带分红的股票价格动态过程给出了QLBS模型在含分红情形下的扩展,第4节利用数值模拟验证了模型在实际市场条件下的优势以及理论正确性。

2. BSM模型

2.1. 经典BSM模型

Black和Scholes [1]在他们的开创性文章中构建了一个对冲组合,即持有一份股票和卖出一份期权并由此不断复制期权。通过分析对冲组合的收益,可以得到期权价格的偏微分方程,并解出期权价格公式。

已知有一标的资产价格 S_t 服从对数正态分布:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (1)$$

现在构造资产组合 Π_t ,包括1单位期权(价值为 C_t)和 $-\delta$ 单位标的资产,则这一资产组合的价值为:

$$\Pi_t = C_t - \delta S_t \quad (2)$$

根据伊藤引理,单位时间内这一资产组合的价值变化为:

$$d\Pi_t = \sigma S_t \left(\frac{\partial C_t}{\partial S_t} - \delta \right) dW_t + \left[\mu S_t \left(\frac{\partial C_t}{\partial S_t} - \delta \right) + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 C_t}{\partial S_t^2} + \frac{\partial C_t}{\partial t} \right] dt \quad (3)$$

现在选择 $\delta = \frac{\partial C_t}{\partial S_t}$,假定无风险利率为 r ,由无套利机会假设和风险中性定价原理,任何无风险资产组合得收益率都等于无风险利率 r ,因此 $d\Pi_t = r\Pi_t dt$,代入上式得:

$$\frac{\partial C_t}{\partial t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 C_t}{\partial S_t^2} + r S_t \frac{\partial C_t}{\partial S_t} - r C_t = 0 \quad (4)$$

这个微分方程的边界条件为当时间 $t = T$ (行权时刻)时,期权价格 C_t 满足 $C_T = \max(Z - S_T; 0)$,其中 Z 是行权价格。根据Hull [9],对于看跌期权,此微分方程的解为:

$$p^{(BS)} = Z e^{-rT} \mathcal{N}(-d_2) - S_0 \mathcal{N}(-d_1) \quad (5)$$

其中，系数 d_1 和 d_2 为：

$$d_1 = \frac{\ln\left(\frac{S_0}{Z}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T} \quad (6)$$

2.2. 考虑分红的BSM模型

假定股息收益率为 q ，则在 dt 时间内持有一份标的资产的股息收入为 $qS_t dt$ 。在 dt 时间内支付股息后，标的资产的价格也将相应地下降相同的幅度 $qS_t dt$ ，此时价格的随机过程为：

$$dS_t = \sigma S_t dW_t + (\mu - q) S_t dt \quad (7)$$

持有数量为 a_t 的标的资产，支付股利后标的资产的下降值为 $qa_t S_t dt$ ，此时投资组合相应地变化为：

$$d\Pi_t = dC_t - \delta dS_t - q\delta S_t dt \quad (8)$$

类似地，我们得到相应连续支付股利的期权所满足的微分方程：

$$\frac{\partial C_t}{\partial t} + \frac{1}{2} \frac{\partial^2 C_t}{\partial S_t^2} \sigma^2 S_t^2 - rC_t + (r - q) \frac{\partial C_t}{\partial S_t} S_t = 0 \quad (9)$$

可以求解得到：

$$p^{(BS')} = Ze^{-rT} \mathcal{N}(-d'_2) - e^{-qT} S_0 \mathcal{N}(-d'_1) \quad (10)$$

其中，系数 d'_1 和 d'_2 为：

$$d'_1 = \frac{\ln\left(\frac{S_0}{Z}\right) + \left(r - q + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}, \quad d'_2 = d'_1 - \sigma\sqrt{T} \quad (11)$$

3. QLBS模型

QLBS模型最早由Halperin [6]提出，它将欧式期权的定价与对冲问题在离散时间节点上重写为一个有限期强化学习问题，是对BSM模型连续时间复制思想的泛化。

在标准设定下，考虑离散时间 $t = 0, 1, \dots, T$ ，时间步长为 Δt 。市场中有一只无分红股票 S_t 和一个无风险资产（银行账户） B_t ，无风险利率为常数 r ，折现因子为 $\gamma = e^{-r\Delta t}$ 。在风险中性测度下，股票价格在一小步内按照几何布朗运动演化，银行账户按无风险利率增长。卖方通过持有股票头寸 a_t 和无风险帐户 B_t 构造自融资投资组合：

$$\Pi_t = a_t S_t + B_t \quad (12)$$

并采用离散时间的Delta 对冲策略，在各个时间步调整 a_t 来对冲标的的价格风险。在强化学习的表述下，QLBS 将期权卖方视为一个RL 智能体：

状态 X_t 取为标的资产的对数，以便消除股票价格动态中的漂移部分；动作 a_t 为在时刻 t 选择的股票头寸；收益 R_t 则为投资组合的贴现交易收益；价值函数采用Q函数形式 $Q_t(x, a)$ ，表示在状态 x 下采取动作 a 时，从 t 到 T 的折现风险调整收益。

在该框架下，卖方的最优定价与对冲问题转化为： $\max_{\pi} Q_0^{\pi}(X_0, a_0)$ ，其中策略 π 给出在各个状态下选择的头寸 a_t 。最优Q-函数 Q_t^* 满足标准的Bellman 递推关系，且对给定状态 X_t ， $Q_t^*(X_t, a_t)$ 对动作 a_t 是二次函数，从而可以求得每一步的最优对冲头寸 $a_t^*(X_t)$ 。在数值实现上，QLBS 使用蒙特卡洛路径模拟和基函数进行回归逼近，对 Q_t^* 和 a_t^* 进行逐步近似，并将 $t = 0$ 时刻最优Q 值的负值视为风险调整后的期权价格。

3.1. QLBS 模型在含分红情形下的扩展

本文主要给出在标的资产按连续股利收益率 q 分红时，QLBS 模型的完整离散时间推导。

3.1.1. 市场设定与含分红的股票增量

记 S_t 为时刻 t 的股价， q 为连续股利收益率。在区间 $[t, t + \Delta t]$ 内，每持有1 股股票获得的股息为 $D_t = qS_t\Delta t$ 。在QLBS 中，我们关心的是“相对于无风险资产”的超额收益。无分红时定义 $\Delta S_t = S_{t+1} - e^{r\Delta t}S_t$ ，即持有一股股票到 $t + 1$ 的价值与将 S_t 投入无风险资产的价值之差。

在含分红情形下，一步内持有一股股票的总价值为“价格+ 股息”： $S_{t+1} + D_t$ 。因此定义含分红的超额收益为：

$$\Delta S_t^{(q)} = S_{t+1} - e^{r\Delta t}S_t + D_t = S_{t+1} - e^{r\Delta t}S_t + qS_t\Delta t \quad (13)$$

当 $q = 0$ 时， $\Delta S_t^{(q)}$ 退化为原来的 ΔS_t 。

3.1.2. 复制组合与自融资约束

考虑投资组合：

$$\Pi_t = a_t S_t + B_t \quad (14)$$

其中 a_t 是在时刻 t 持有的股票头寸， B_t 为无风险账户头寸。

初始时刻 t 时，组合价值为 Π_t 。到 $t + 1$ 前一刻，银行账户获得利息 $e^{r\Delta t}B_t$ ，同时股票支付股利 $a_t D_t$ ，因此 $B_{t+1}^- = e^{r\Delta t}B_t + a_t D_t$ 。此时的组合价值为：

$$\Pi_{t+1}^- = a_t S_{t+1} + B_{t+1}^- \quad (15)$$

在时刻 $t + 1$ 调整头寸到 a_{t+1} ，自融资条件要求调仓前后组合价值一致，可得时刻 $t + 1$ 的组合价值：

$$\begin{aligned}
\Pi_{t+1} &= a_{t+1}S_{t+1} + B_{t+1} \\
&= a_t S_{t+1} + B_{t+1} \\
&= a_t S_{t+1} + e^{r\Delta t} B_t + a_t D_t
\end{aligned} \tag{16}$$

从而:

$$B_t = e^{-r\Delta t} (\Pi_{t+1} - a_t(S_{t+1} + D_t)) = \gamma(\Pi_{t+1} - a_t(S_{t+1} + D_t)) \tag{17}$$

其中 $\gamma = e^{-r\Delta t}$, 代回 $\Pi_t = a_t S_t + B_t$, 得到:

$$\begin{aligned}
\Pi_t &= a_t S_t + \gamma(\Pi_{t+1} - a_t(S_{t+1} + D_t)) \\
&= \gamma \Pi_{t+1} - \gamma a_t (S_{t+1} + D_t - e^{r\Delta t} S_t) \\
&= \gamma(\Pi_{t+1} - a_t \Delta S_t^{(q)})
\end{aligned} \tag{18}$$

终端条件为 $\Pi_T = H(S_T) = \max(Z - S_T, 0)$ 。

3.1.3. 最优动作价值函数和奖励函数

在状态变量的选择上, 按照QLBS原文中的做法, 我们引入时间齐次的状态变量 X_t ,但在含分红情形下, 对 X_t 的定义改为:

$$X_t = -\left((\mu - q) - \frac{\sigma^2}{2}\right)t + \ln S_t \tag{19}$$

则有:

$$S_t = e^{X_t + \left((\mu - q) - \frac{\sigma^2}{2}\right)t} \tag{20}$$

并且在连续时间极限下:

$$dX_t = \sigma dW_t \tag{21}$$

即 X_t 是一个鞅, 并且是时间均匀的, 不像股票价格 S_t 有漂移。

我们将公平期权价格 \hat{C}_t 定义为对冲投资组合 Π_t 的期望值:

$$\hat{C}_t = \mathbb{E}_t [\Pi_t | \mathcal{F}_t] \tag{22}$$

实际的公允风险调整价格应该由卖方的风险偏好决定, 故应该纳入由对冲投资组合的贴现方差给出的风险溢价, 并按风险厌恶参数 λ 进行缩放:

$$\begin{aligned}
C_0(S, a) &= \hat{C}_0 + \lambda \mathbb{E}_0 \left[\sum_{t=0}^T e^{-rt} \text{Var} [\Pi_t | \mathcal{F}_t] \mid S_0 = S, a_0 = a \right] \\
&= \mathbb{E}_0 \left[\Pi_0 + \lambda \sum_{t=0}^T e^{-rt} \text{Var} [\Pi_t | \mathcal{F}_t] \mid S_0 = S, a_0 = a \right]
\end{aligned} \tag{23}$$

为了有效控制风险, 并确保卖方在未来不会因对冲组合的失效而面临巨大亏损, 期权卖方需要最小化期权价格 $C_0(S, a)$ 。同时我们注意到, 在RL框架中, 公允期权价格的最小化问题等价于最

大化该时刻的动作价值函数。因此可以表示为 $Q_t = -C_t$ 。此时，在状态变量 X_t 和策略 π 下的动作价值函数可以表示为：

$$\begin{aligned} Q_t^\pi(x, a) &= \mathbb{E}_t \left[-\Pi_t(X_t) - \lambda \sum_{t'=t}^T e^{-r(t'-t)} \text{Var} [\Pi_{t'}(X_{t'}) | \mathcal{F}_{t'}] \middle| \mathcal{F}_t \right] \\ &= \mathbb{E}_t \left[-\Pi_t(X_t) - \lambda \text{Var} [\Pi_t] - \lambda \sum_{t'=t+1}^T e^{-r(t'-t)} \text{Var} [\Pi_{t'}(X_{t'}) | \mathcal{F}_{t'}] \middle| \mathcal{F}_t \right] \\ &= \mathbb{E}_t [-\Pi_t(X_t) - \lambda \text{Var} (\Pi_t(X_t)) + \gamma (Q_{t+1}^\pi(X_{t+1}, a_{t+1}) + \mathbb{E}_{t+1} [\Pi_{t+1}(X_{t+1})])] \end{aligned} \quad (24)$$

根据贝尔曼期望方程，我们有：

$$Q^\pi(x, a) = \mathbb{E}^\pi [R_{t+1} + \gamma Q^\pi(x' = X_{t+1}, a' = A_{t+1}) | x = X_t, a = A_t] \quad (25)$$

由上面两个式子， $t \in \{0, 1, \dots, T-1\}$ 时刻的回报 R_t 可表示为：

$$\begin{aligned} R_t(X_t, a_t, X_{t+1}) &= \gamma \Pi_{t+1}(X_{t+1}) - \Pi_t(X_t) - \lambda \text{Var} [\Pi_t | \mathcal{F}_t] \\ &= \gamma \Pi_{t+1}(X_{t+1}) - \gamma (\Pi_{t+1} - a_t \Delta S_t^{(q)}) - \lambda \text{Var} [\Pi_t | \mathcal{F}_t] \\ &= \gamma a_t \Delta S_t^{(q)} - \lambda \text{Var} [\Pi_t | \mathcal{F}_t] \end{aligned} \quad (26)$$

对于最后一个时间步 $t = T$, 回报为 $R_T = -\lambda \text{Var} [\Pi_T]$ 。

再结合回报 R_T 和动作价值函数得出最优动作价值函数：

$$\begin{aligned} Q_t^{*,(q)}(X_t, a_t) &= \gamma \mathbb{E}_t [Q_{t+1}^{*,(q)}(X_{t+1}, a_{t+1}^{*,(q)} + a_t \Delta S_t^{(q)})] - \lambda \text{Var} [\Pi_t(X_t)] \\ &= \gamma \mathbb{E}_t [Q_{t+1}^{*,(q)}(X_{t+1}, a_{t+1}^{*,(q)} + a_t \Delta S_t^{(q)})] - \lambda \gamma^2 \mathbb{E}_t [\hat{\Pi}_{t+1}^2 - 2a_t \hat{\Pi}_{t+1} \Delta \hat{S}_t^{(q)} + a_t^2 (\Delta \hat{S}_t^{(q)})^2] \\ &= \gamma \mathbb{E}_t [Q_{t+1}^{*,(q)}(X_{t+1}, a_{t+1}^{*,(q)} + a_t \Delta S_t^{(q)})] - \lambda \gamma^2 \mathbb{E}_t \left[\left(\hat{\Pi}_{t+1} - a_t (\Delta \hat{S}_t^{(q)}) \right)^2 \right] \end{aligned} \quad (27)$$

其中， $\hat{\Pi}_{t+1} := \Pi_{t+1} - \mathbb{E}_t(\Pi_{t+1})$ ， $\Delta \hat{S}_t^{(q)} := \Delta S_t^{(q)} - \mathbb{E}_t(\Delta S_t^{(q)})$ 。

从上式中可以看出 $Q_t^{*,(q)}(X_t, a_t)$ 是关于 a_t 的二次函数，因此只需对 $-Q_t^{*,(q)}(X_t, a_t)$ 对 a_t 求导并令其为零，可得含分红情形下的最优对冲解析解：

$$a_t^{*,(q)}(X_t) = \frac{\mathbb{E}_t [\Delta \hat{S}_t^{(q)} \hat{\Pi}_{t+1}] + \frac{1}{2\gamma\lambda} \mathbb{E}_t [\Delta S_t^{(q)}]}{\mathbb{E}_t [(\Delta \hat{S}_t^{(q)})^2]} \quad (28)$$

与无分红情形相比，上式将 ΔS_t 、 $\Delta \hat{S}_t$ 替换为含分红的 $\Delta S_t^{(q)}$ 、 $\Delta \hat{S}_t^{(q)}$ ，在分红率 $q = 0$ 时，最优对冲解析解与原始QLBS模型完全一致，严格退化为原始QLBS框架。

3.1.4. 函数逼近与含分红QLBS 期权价格

在有模型 (Model-Based) QLBS 中, 可采用线性基函数逼近最优对冲与最优 Q -函数:

$$a_t^{*,(q)}(X_t) \approx \sum_{n=1}^N \phi_{n,t}^{(q)} \Phi_n(X_t) \tag{29}$$

$$Q_t^{*,(q)}(X_t, a_t^{*,(q)}(X_t)) \approx \sum_{n=1}^N \omega_{n,t}^{(q)} \Phi_n(X_t) \tag{30}$$

其中 $\{\Phi_n(\cdot)\}_{n=1}^N$ 为预先选定的基函数族, $\phi_{n,t}^{(q)}$ 、 $\omega_{n,t}^{(q)}$ 与时间有关并且在时间 $t = T - 1, \dots, 0$ 的时间内向后递归计算得到。与无分红情形相比, 唯一的区别是所有涉及 ΔS_t 的样本量与统计量都替换为 $\Delta S_t^{(q)}$ 。

对于基函数的选择问题, Grau [10]发现不同的基函数对定价精度和计算效率有显著影响, 主要取决于问题的维度和对光滑性的要求。对于欧式期权的定价与对冲问题上, B样条插值 (B-splines) 能够提供高效的局部逼近, 并且在逼近时具有较好的光滑性。

最终, 含分红QLBS 在 $t = 0$ 时的“风险调整后期权价格”可写为:

$$C_0^{\text{QLBS},(q)} = -\mathbb{E}[Q_0^{*,(q)}(X_0, a_0^{*,(q)}(X_0))] \approx -\frac{1}{K} \sum_{k=1}^K Q_0^{*,(q)}(X_0^{(k)}, a_0^{*,(q)}(X_0^{(k)})) \tag{31}$$

4. 数值实验结果与分析

4.1. 参数设置与路径模拟

为了验证QLBS^(q)模型在实际市场条件下的理论正确性, 本节将展示带红利因素的BSM与QLBS^(q)的期权价格和最优对冲并将它们进行比较。参考Stoiljkovic在文献 [8]中所使用的参数设置, 我们的模型参数见表 1所示:

Table 1. Parameter settings

表 1. 参数设置

初始股价	股票漂移率	股票波动率	无风险利率	行权价	时间	蒙特卡洛路径数	B样条个数	B样条	股息率
S_0	μ	σ	r	Z	步数	K	N	阶数	q
100	0.03	0.15	0.05	100	24	10 000	12	4	0.02

表 1列出了用于模型数值实现的参数。期权到期日设置为1年, 风险厌恶参数 $\lambda = 0.001$ 。在风险中性测度下, 折现后的累计收益过程必然是一个鞅。对于一只支付连续股息收益率 q 的股票, 其预期收益必须等于无风险利率 r 。因此, 价格过程的漂移率 μ 被调整为 $r - q$, 以使预期总收益保持在 r 。此外, 根据Halperin [6] 和Dixon [11], 我们通过设置 $\frac{1}{2\gamma\lambda} = 0$ 来调整最优对冲, 以此排除风险厌恶对期权定价和对冲问题的影响。满足交易者对风险中性的假设。

根据上述设置的参数，我们采用蒙特卡洛模拟来描述未来标的资产价格的变化情况。在蒙特卡洛模拟中，时间是离散的，通过选择合适的时间步长 $\Delta t = T/n$,可以将连续时间过程离散化，在每个时间步长上，资产价格的变化通过对几何布朗运动进行模拟来实现。通过这种方式，沿着每个时间步长生成资产价格路径并重复这个过程，我们可以得到大量不同的资产价格轨迹。参数按照上述设置得到的蒙特卡洛模拟路径见表 2所示：

Table 2. Stock price path simulation results

表 2. 股价路径模拟结果

	0	1	2	...	22	23	24
1	100	115.58	119.62	...	74.92	84.48	86.03
2	100	91.87	86.56	...	43.49	40.45	40.49
...
9999	100	104.43	113.29	...	100.47	105.57	112.25
10000	100	100.86	84.54	...	117.34	120.87	123.52

在本文中，由于QLBS^(q)模型设置的状态变量为单一维度且期权定价的应用需要光滑逼近，我们选择使用B-splines 作为基函数对最优对冲和最优动作价值函数进行逼近，图 1展示了样条数量为12，阶数为4的B-splines：

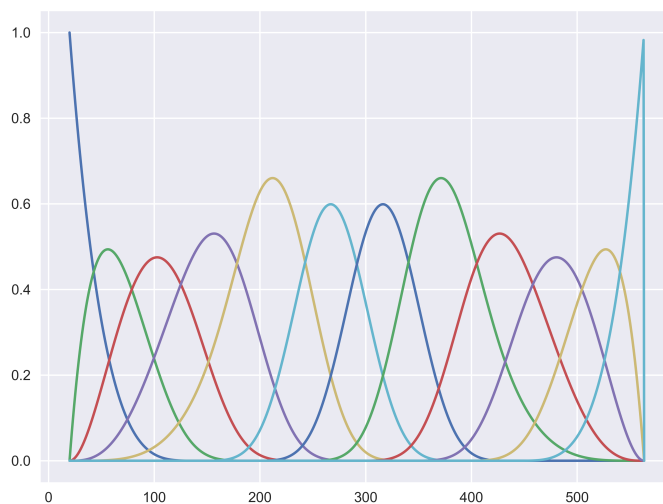


Figure 1. The graph of the fourth-order basis functions of the B-spline function ($N = 12$)

图 1. B样条函数 ($N = 12$) 的四阶基函数图像

4.2. 数值实验

4.2.1. 模型合理性检验

在第一组实验中，我们将股息收益率设为 $q = 0$ ，为了验证模型的正确性，我们将QLBS^(q)在不同行权价 $Z \in [80, 120]$ 下给出的欧式看跌期权价格，与Black-Scholes无分红模型解析解进行比

较。表 3和图 2 给出了两种方法的期权价格和对冲的结果：

Table 3. QLBS^(q) vs BSM put option prices and hedging ($t = 0, q = 0$)

表 3. QLBS^(q) vs BSM 看跌期权价格与对冲 ($t = 0, q = 0$)

行权价Z	BSM 价格	QLBS ^(q) 价格	价格相对误差(%)	BSM 对冲	QLBS ^(q) 对冲	对冲相对误差(%)
80	0.25	0.23	8.00	-0.04	-0.03	25.00
90	1.40	1.36	2.86	-0.16	-0.14	12.50
100	4.53	4.47	1.32	-0.39	-0.35	10.26
110	10.13	10.07	0.59	-0.64	-0.59	7.81
120	17.76	17.72	0.23	-0.83	-0.77	7.23

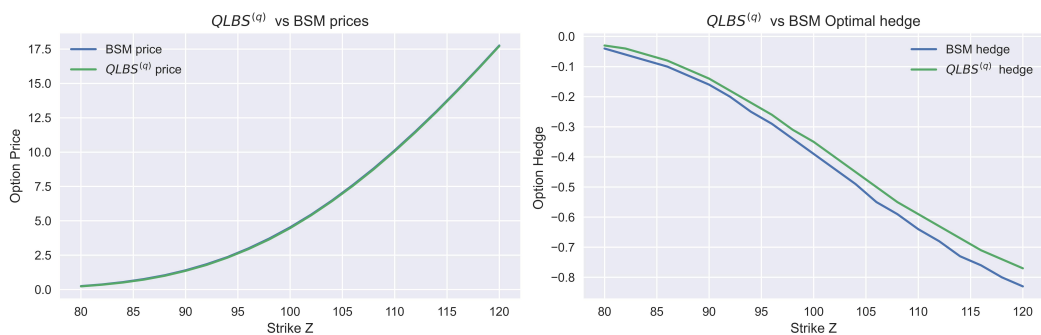


Figure 2. QLBS^(q) vs BSM put option prices and hedging ($t = 0, q = 0$)

图 2. QLBS^(q) vs BSM 看跌期权价格与对冲 ($t = 0, q = 0$)

图 2左图展示了不同行权价时得到的期权价格曲线，可以看到QLBS^(q)定价与BSM解析解几乎重合。随着行权值的上升，看跌期权价格也会随之增加，对冲比率随之降低。表 3给出了对应的数值结果，对于平值与实值期权 ($Z \geq 100$)，QLBS^(q)的价格相对误差均在2%以内。在深度虚值期权 $Z = 80$ 时，由于期权价格值较小，导致相对误差被放大。图 2右图比较了时刻 $t = 0$ 时QLBS^(q)学习得到的最优对冲 a_0^* 与BSM 模型的解析Delta。可以看到两条对冲曲线随行权价的变化趋势几乎一致，对于深度虚值期权，QLBS^(q)和BSM 对冲比率几乎相同。然而随着期权的价内深度加深，对冲比率开始出现更加明显的差异。总体来看，在无分红情形下，本文实现的QLBS^(q)模型可以准确复现经典BSM模型的定价结果，同时也能较好地逼近最优Delta对冲策略。

4.2.2. 定价精度验证

在第二个实验中，我们固定行权价 $Z = 100$ ，考察不同股利收益率 $q \in [0\%5\%]$ 对期权定价的影响。图 3和表 4展示了QLBS^(q)模型和含分红BSM 模型在该参数区间内的欧式看跌期权价格。

可以看到，股息的纳入降低了标的资产的漂移率，导致看跌期权价格上涨。QLBS^(q)模型有效地捕捉了这种行为，这与理论预期一致。随着股利收益率 q 的上升，期权价格单调增加，并且QLBS^(q)曲线在整个区间几乎与解析的BSM曲线重合。表 4给出了若干股利收益率下的数值比较， $q = 0$ 的情形已在实验1中讨论，仅列出 $q > 0$ 的结果。对于含分红的情形，QLBS^(q)与BSM的价格误差均在1%左右。对于最优对冲头寸 a_0^* ，两条曲线均随 q 单调下降，表明在股息收益率较高

Table 4. QLBS^(q) vs BSM put option prices and hedging $q \in [0\%, 5\%]$

表 4. QLBS^(q) vs BSM 看跌期权价格与对冲 ($q \in [0\%, 5\%]$)

股息收益率 q (%)	BSM 价格	QLBS ^(q) 价格	价格相对误差(%)	BSM 对冲	QLBS ^(q) 对冲	对冲相对误差(%)
1000	5.36	5.42	1.11	-0.43	-0.33	23.626
4000	5.36	5.31	0.93	-0.43	-0.37	13.95
7000	5.36	5.33	0.56	-0.43	-0.39	9.30
10000	5.36	5.34	0.37	-0.43	-0.40	6.98
13000	5.36	5.33	0.56	-0.43	-0.40	6.98

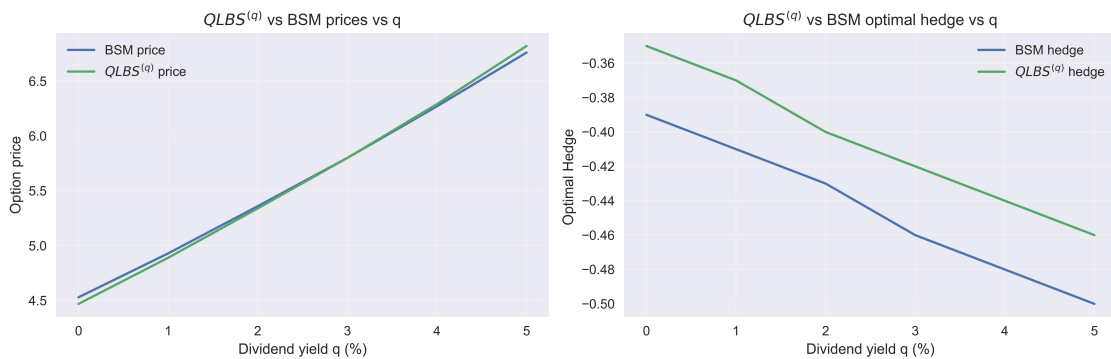


Figure 3. QLBS^(q) vs BSM put option prices and hedging, $q \in [0\%, 5\%]$

图 3. QLBS^(q) vs BSM 看跌期权价格与对冲, $q \in [0\%, 5\%]$

时，看跌期权更趋近于实值，需要持有更多空头标的来进行对冲。与解析解相比，QLBS^(q)的对冲头寸略微偏向零，对冲相对误差大约维持在7%-10%左右。可能的原因如下：QLBS^(q)采用离散时间框架，而BSM Delta是连续时间下的解析对冲，当期权存续期被离散为有限个时间步时，会不可避免地产生对连续时间最优对冲的近似误差。其次，最优对冲头寸的估计依赖于条件协方差和条件方差等样本统计量，而本文蒙特卡洛模拟路径数为10000条，对对冲头寸这类对局部统计误差更敏感的量而言，仍可能带来一定的系统偏移。总的来说，在引入连续股利后，QLBS^(q)模型在定价与对冲层面均能稳定逼近含分红的BSM基准。说明本文引入股利因素的QLBS^(q)在定价层面是有效且稳定的。

4.2.3. 对冲PnL分布实验

在实验3中，我们采用与前文相同的市场参数，并在同一组路径上分别应用含分红BSM Delta与QLBS^(q)的最优对冲策略。对每条路径，我们根据递推式：

$$V_{t+\Delta t} = e^{r\Delta t} (V_t - a_t S_t) + a_t (S_{t+\Delta t} + q S_t \Delta t) \tag{32}$$

计算对冲组合的终端价值 V_t ，将对冲PnL定义为：

$$\varepsilon = V_T - \max(Z - S_T, 0) \tag{33}$$

并统计其分布特征。

图 4左图给出了两种策略的终端对冲误差直方图，右图展示了对冲误差的标准差和95% 分位数等风险指标。



Figure 4. Hedging PnL distribution and risk indicators

图 4. 对冲PnL 分布及风险指标

可以看到，Delta 对冲与QLBS^(q)对冲的误差分布均以0为中心呈近似对称的单峰形状，表明两种对冲的PnL基本无偏。并且两条分布几乎重合，右图展示了对冲误差的标准差和95% 分位数，QLBS^(q)的风险指标略高于解析Delta，综上，QLBS^(q)在动态对冲层面的表现与含分红BSM Delta接近，可视为对该基准的高精度数值逼近方法。

4.2.4. 参数稳健性分析

最后，为验证本文中几个关键超参数的稳健性，我们在固定连续股利率 $q = 0.02$ 的条件下，分别考察了蒙特卡洛模拟路径数 K 、时间步数对定价误差和对冲误差的影响。结果如表 5和表 6显示：

Table 5. Volatility model coefficients

表 5. 蒙特卡洛路径数的稳健性分析

路径数 K	BSM价格	QLBS价格	价格误差(%)	BSM Delta	QLBS对冲 a_0^*	对冲误差(%)
1000	5.36	5.42	1.11	-0.43	-0.33	23.626
4000	5.36	5.31	0.93	-0.43	-0.37	13.95
7000	5.36	5.33	0.56	-0.43	-0.39	9.30
10000	5.36	5.34	0.37	-0.43	-0.40	6.98
13000	5.36	5.33	0.56	-0.43	-0.40	6.98

上述两组稳健性分析表明，QLBS^(q)数值结果对超参数设定具有一定依赖性，蒙特卡洛路径数的增加有助于降低统计误差，尤其能显著改善对冲头寸的稳定性；时间步数的增加虽然理论上可减小离散时间近似误差，但在有限样本条件下会同时放大噪声并加剧误差。本文所采用的基准参数能够在计算成本可控的前提下保持较高的定价精度，但若以对冲精度为主要目标，则适当增加蒙特卡洛路径数可能是更优的选择。

Table 6. Robustness analysis of timesteps**表 6.** 时间步数的稳健性分析

时间步数	BSM价格	QLBS价格	价格误差(%)	BSM Delta	QLBS对冲 a_0^*	对冲误差(%)
12	5.36	5.32	0.75	-0.43	-0.41	5.06
24	5.36	5.34	0.37	-0.43	-0.40	6.98
36	5.36	5.33	0.56	-0.43	-0.40	7.93
48	5.36	5.32	0.75	-0.43	-0.38	13.22
60	5.36	5.30	1.12	-0.43	-0.39	10.39

5. 结论

本研究通过将股利支付纳入传统的QLBS模型，验证了本文实现的QLBS^(a)模型可以精准且稳定地逼近Black-Scholes模型的定价结果和最优对冲策略，验证了施加股利因素的QLBS模型的可能性及有效性，更好地反映实际市场情况，为投资者提供更有价值的决策支持。

参考文献

- [1] Black, F. and Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81**, 637-654. <https://doi.org/10.1086/260062>
- [2] Merton, R.C., et al. (1971) Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, **4**, 141-183. <https://doi.org/10.2307/3003143>
- [3] Krausz, J. (1985) Option Parameter Analysis and Market Efficiency Tests: A Simultaneous Solution Approach. *Applied Economics*, **17**, 885-896. <https://doi.org/10.1080/00036848500000042>
- [4] Broadie, M. and Detemple, J. (1995) American Capped Call Options on Dividend-Paying Assets. *Review of Financial Studies*, **8**, 161-191. <https://doi.org/10.1093/rfs/8.1.161>
- [5] Sutton, R.S., Barto, A.G., et al. (1998) Reinforcement Learning: An Introduction, Volume 1. MIT Press.
- [6] Halperin, I. (2017) QLBS: Q-Learner in the Black-Scholes (-Merton) Worlds. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3087076>
- [7] Halperin, I. (2019) The QLBS Q-Learner Goes NuQLear: Fitted Q Iteration, Inverse RL, and Option Portfolios. *Quantitative Finance*, **19**, 1543-1553. <https://doi.org/10.1080/14697688.2019.1622302>
- [8] Stoiljkovic, Z. (2025) Advanced Option Pricing and Hedging with Q-Learning: Performance Evaluation of the QLBS Algorithm. *The Journal of Derivatives*, **32**, 48-79. <https://doi.org/10.3905/jod.2025.1.222>

-
- [9] Hull, J.C. and Basu, S. (2016) Options, Futures, and Other Derivatives. Pearson Education India.
- [10] Grau, A.J. (2008) Applications of Least-Squares Regressions to Pricing and Hedging of Financial Derivatives. PhD Thesis, Technische Universität München.
- [11] Dixon, M.F., Halperin, I., Bilokon, P., *et al.* (2020) Machine Learning in Finance, Volume 1170. Springer.