



Article

How Much Do Negative Probabilities Matter in Option Pricing?: A Case of a Lattice-Based Approach for Stochastic Volatility Models

Chung-Li Tseng ^{1,*} , Daniel Wei-Chung Miao ², San-Lin Chung ³ and Pai-Ta Shih ³

¹ UNSW Business School, The University of New South Wales, Sydney, NSW 2052, Australia

² Graduate Institute of Finance, National Taiwan University of Science and Technology, Taipei 106335, Taiwan; miao@mail.ntust.edu.tw

³ Department of Finance, National Taiwan University, Taipei 10617, Taiwan; chungsl@management.ntu.edu.tw (S.-L.C.); ptshih@management.ntu.edu.tw (P.-T.S.)

* Correspondence: c.tseng@unsw.edu.au

Abstract: In this paper, we focus on two-factor lattices for general diffusion processes with state-dependent volatilities. Although it is common knowledge that branching probabilities must be between zero and one in a lattice, few methods can guarantee *lattice feasibility*, referring to the property that all branching probabilities at all nodes in all stages of a lattice are legitimate. Some practitioners have argued that negative probabilities are not necessarily ‘bad’ and may be further exploited. A theoretical framework of lattice feasibility is developed in this paper, which is used to investigate how negative probabilities may impact option pricing in a lattice approach. It is shown in this paper that lattice feasibility can be achieved by adjusting a lattice’s configuration (e.g., grid sizes and jump patterns). Using this framework as a benchmark, we find that the values of out-of-the-money options are most affected by negative probabilities, followed by in-the-money options and at-the-money options. Since legitimate branching probabilities may not be unique, we use an optimization approach to find branching probabilities that are not only legitimate but also can best fit the probability distribution of the underlying variables. Extensive numerical tests show that this optimized lattice model is robust for financial option valuations.

Keywords: finance; trinomial tree; two-factor model; stochastic volatility; lattice feasibility



Citation: Tseng, Chung-Li, Daniel Wei-Chung Miao, San-Lin Chung, and Pai-Ta Shih. 2021. How Much Do Negative Probabilities Matter in Option Pricing?: A Case of a Lattice-Based Approach for Stochastic Volatility Models. *Journal of Risk and Financial Management* 14: 241. <https://doi.org/10.3390/jrfm14060241>

Academic Editor: Robert Brooks

Received: 12 April 2021

Accepted: 23 May 2021

Published: 30 May 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The lattice (or tree) approach is a popular one for valuing derivative securities, as it is normally simple to implement and has an intuitive appeal. The lattice approach involves discrete approximation to the diffusion processes followed by the underlying variables. It is especially useful for valuing American options where early exercise is possible. Since its introduction by Cox et al. (1979), the lattice approach has undergone several extensions in the past few decades to accommodate increasingly complex derivative valuations. To name a few, those significant models include Rendleman and Bartter (1979), Boyle (1986, 1988), (H&W) Hull and White (1988, 1990, 1993, 1994), Chung and Shih (2007), Beliaeva and Nawalkha (2010), and Akyildirim et al. (2014).

In a lattice, each link (branch) connecting two lattice nodes at two consecutive time periods is associated with a branching probability. A legitimate branching probability must be between zero and one. Researchers know and follow this basic rule in developing lattice-based methods. However, to the best of our knowledge, few methods can guarantee *lattice feasibility*, referring to the property that all branching probabilities at all nodes in all stages of a lattice are legitimate. With lattice feasibility, a lattice constructs a discrete time financial market that is arbitrage free. It is well known that lattice feasibility is easier to achieve when there is only one underlying variable, while two-factor lattice feasibility

is harder to meet, especially when the correlation between two underlying uncertainties is high.

The term lattice feasibility was coined by [Tseng and Lin \(2007\)](#). The authors employed the trinomial lattice proposed by H&W ([Hull and White 1990](#)) to value real options involving two underlying correlated uncertainties, each with a constant volatility. They found that each lattice configuration implies a maximum correlation of the two underlying variables that the lattice can approximate without incurring negative probabilities, and this maximum correlation may be enhanced by varying the size of its lattice grids. After optimizing the lattice configuration, [Tseng and Lin \(2007\)](#) also showed that the trinomial lattice proposed by H&W ([Hull and White 1990](#)) cannot accommodate a correlation beyond $4/\sqrt{35} \approx 0.676$ without incurring negative probabilities. The authors further showed that the popular two-factor interest rate tree proposed by H&W ([Hull and White 1994](#)) for valuing interest rate derivatives can only guarantee lattice feasibility when the correlation is no greater than 0.2. This means that negative probabilities may occur far more often than we know in using lattices to value derivatives in real practice.

Since it is not unusual to encounter negative probabilities when using lattices for option pricing, some practitioners have argued that negative probabilities may not necessarily be 'bad' and may be further exploited (e.g., [Burgin and Meissner 2012](#); [Haug 2007](#)). Consider a price node in a trinomial tree, where the sum of three branching probabilities must be equal to one. Given an abnormality where one branching probability becomes negative or exceeds unity, the other two probabilities must be adjusted accordingly to offset the abnormality. However, the expected payoff at this price node may reveal no sign of abnormality. From this perspective, it is not surprising that some practitioners have reported that some finite difference/finite element models can still produce stable and consistent outputs even with negative probabilities ([Zvan et al. 2001](#)). How much does it really matter if one allows negative probabilities in a lattice for option pricing?

To investigate the impact of negative probabilities in option valuations, we focus on using a two-factor lattice to represent general diffusion processes such as the Heston stochastic volatility (SV) model ([Heston 1993](#)). In the Heston model, the dynamics of the volatility process is assumed to follow the CIR process ([Cox et al. 1985](#)) used to describe the interest rate dynamics. The analytical tractability of the CIR process leads to explicit solutions for some bond pricing problems (e.g., [Kouritzin 2000](#); [Maghsoodi 1996](#); among others). When the CIR process is incorporated as the second dimension of the Heston model, the resultant two-factor lattice is more general and is far from trivial. To observe the impact that comes from negative probabilities, we need to develop a lattice model that can guarantee lattice feasibility and can be used as a benchmark. Under the same lattice framework, with every parameter fixed but branching probabilities, we can then observe how negative probabilities influence valuations.

An important alternative to the lattice method for the options pricing under the Heston model is the Monte Carlo (MC) method. Since the MC method generates stochastic paths, not lattices, to model the evolution of the underlying uncertainties, there is no issue of lattice feasibility or negative probability. However, the generations of the stochastic paths under the Heston model are not straightforward, and there have been a number of discussions on this issue. Examples along this line include [Broadie and Kaya \(2006\)](#) and [Andersen \(2007\)](#). When the MC method is applied to the pricing of American options, the Least-Squares Monte Carlo (LSMC) method is a standard approach proposed by [Longstaff and Schwartz \(2001\)](#). Though a lot of progresses have been made using the MC method to price American options, one also needs to resort to its variations, such as different techniques on resampling or branching (e.g., see a recent paper by [Kouritzin and Mackay 2020](#)).

Researchers have proposed lattices for stochastic volatility models. Recent papers include ([Akyildirim et al. 2014](#); [Beliaeva and Nawalkha 2010](#); [Costabile et al. 2012](#); [Ruckdeschel et al. 2013](#)). All lattice approaches consider matching two conditional marginal moments for each underlying variable at all nodes, and the correlation is dealt with either by matching the cross moment of the variables or using variable transformation to decorre-

late them. Special attention must be paid to avoid negative branching probabilities that are more likely to occur when the correlation is high. One popular approach is to truncate branching probabilities that are negative or exceed unity to bring them to be between zero and one. While truncating branching probabilities may not exactly match the moments, [Akyildirim et al. \(2014\)](#) show that in their approach the matching error may be negligible and prove the convergence of their approach to the underlying processes. In this paper, we take the standard approach by matching the two marginal moments and the cross moment of the two underlying variables. We show that branching probabilities can be guaranteed to be between zero and one by adjusting the configuration of the lattice for a given (fixed) time step, and no probability truncation is needed.

To manage stochastic volatilities, we extend the lattice parameters from the grid size to include a jump size. With this change, the lattice configuration can be optimized to guarantee lattice feasibility even if both state variables are highly correlated with the correlation close to one. As will be shown later, this newly introduced parameter, the jump size, has the effect of refining the grid size. As opposed to traditional lattice approaches which perform lattice refinement on time space, our method can also perform refinement on the state space of the underlying variables even when the time step is not especially small. The consequence is better fitting of the underlying processes and faster convergence.

Numerical tests show that lattice feasibility has a direct impact on option pricing. We find that the values of out-of-the-money (OTM) options are most affected by negative probabilities, followed by in-the-money (ITM) options and at-the-money (ATM) options. Although negative probabilities matter less in some situations, the resulted distortion of the underlying probability distribution is in general hard to predict and exploit. Despite the importance of lattice feasibility, our numerical tests also show that lattice feasibility alone may not be sufficient to guarantee accurate valuation, especially when the time step of the lattice is not especially small. Since legitimate branching probabilities may not be unique, we propose an optimization approach to find branching probabilities that are not only legitimate but also can best fit the probability distribution of the underlying variables.

The rest of this paper is structured as follows. To lay the foundation of lattice feasibility for approximating two general, correlated diffusion processes, a general one-factor lattice model is first considered in Section 2, with the CIR model used as an example to show in detail how the lattice can be constructed. Section 3 considers the lattice for two general and correlated diffusion processes including the Heston SV model ([Heston 1993](#)), and derives the lattice feasibility conditions. We analyze the impact of lattice feasibility on options valuation in Section 4 and conduct extensive numerical tests, including pricing European options and American options in Section 5. This paper concludes in Section 6. All proofs of propositions and theorems are given in the Appendix A of this paper.

2. General One-Factor Trinomial Lattice

We consider the following general Ito process:

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t, \tag{1}$$

where W_t is a Wiener process and the volatility $\sigma(\cdot)$ is a state-dependent function. In this paper, we assume that $\sigma(\cdot)$ is lower bounded by a constant $\sigma^{\min} > 0$, which is a small number close to zero. That is,

$$\sigma(y) \geq \sigma^{\min} > 0, \forall y. \tag{2}$$

For any process whose volatility function does not satisfy (2), denoted as

$$dY_t = \mu(Y_t)dt + \alpha(Y_t)dW_t, \tag{3a}$$

where $\alpha(y) \geq 0$ (e.g., $\alpha(y) = \xi\sqrt{y}$ in the CIR model), one can work with the following counterpart:

$$dY_t = \mu(Y_t)dt + \max(\alpha(Y_t), \sigma^{\min})dW_t, \tag{3b}$$

whose volatility function satisfies (2). In (3b), the volatility is assumed to be bounded below by σ^{\min} for the convenience of subsequent treatments. For a sufficiently small σ^{\min} , the effect brought about by this lower bound is insignificant and the difference between the two models is negligible, as long as the drift $\mu(\cdot)$ is nonzero when the volatility is close to zero with the process on the brink of being absorbed. How to determine the lower bound σ^{\min} in the setting of (3b) will be addressed later.

We will propose a trinomial lattice to approximate (1) based on H&W (Hull and White 1993). In their model, the time horizon is divided into intervals of equal length Δt , and the process can only take on values that are multiples of Δy . At time t , a typical lattice node $Y_t = y$ branches to nodes $y + (k + h)\Delta y$, $y + k\Delta y$, and $y + (k - h)\Delta y$ at the next stage with respective branching probabilities p_u , p_m , and p_d , where the (text) subscripts represent upward, middle, and downward branches, respectively. As will be shown later, the jump size h depends on y and, therefore, the lattice may not recombine at all nodes.

The lattice size $\Delta y = c\sigma^s\sqrt{\Delta t}$ is predetermined using the same method as when volatility is constant, where c and σ^s are constant. Since $\sigma(y)$ is a function, σ^s is a surrogate constant volatility, which is set to be a small number no greater than σ^{\min} :

$$0 < \sigma^s \leq \sigma^{\min}. \tag{4}$$

The branching factor k is chosen such that $k\Delta y$ (middle branch) approximates the expected price deviation $\mu(y)\Delta t$. That is, k is the nearest integer of $\mu(y)\Delta t / \Delta y$:

$$k(y) \equiv \left\lfloor \mu(y)\Delta t / \Delta y + 0.5 \right\rfloor, \tag{5}$$

where $\lfloor \cdot \rfloor$ is the floor function that maps to the nearest integer less than or equal to the operand. Let the mismatch between $\mu(y)\Delta t / \Delta y$ and $k(y)$ in (5) be denoted by ϵ , defined as follows:

$$\epsilon(y) \equiv \mu(y)\Delta t / \Delta y - k(y). \tag{6}$$

Apparently from (5),

$$|\epsilon| \leq 0.5, \forall y. \tag{7}$$

The interpretation of (7) is that the middle branch may miss the expected price deviation by no more than $0.5\Delta y$. Note that both ϵ and k are functions of y and t , and may vary from node to node.

The jump size h is determined such that $h\sigma^s$ approximates $\sigma(y)$. Likewise, let h be the nearest integer of $\sigma(y) / \sigma^s$:

$$x(y) \equiv \sigma(y) / \sigma^s, \tag{8}$$

$$h(y) \equiv \left\lfloor \sigma(y) / \sigma^s + 0.5 \right\rfloor = \left\lfloor x(y) + 0.5 \right\rfloor. \tag{9}$$

Depending on how small σ^s is chosen, there is a minimum value of h incurred in (9) and is denoted by \underline{h} :

$$\underline{h} = \min_y h(y) = \left\lfloor \sigma^{\min} / \sigma^s + 0.5 \right\rfloor. \tag{10}$$

Given σ^{\min} , one can either choose a σ^s to determine the integer \underline{h} , as in (10); or vice versa. In our design, one determines \underline{h} first, followed by σ^s using the following formula:

$$\sigma^s = \frac{\sigma^{\min}}{\max(\underline{h} - 0.5, 1)} = \begin{cases} \sigma^{\min}, & \text{if } \underline{h} = 1, \\ \sigma^{\min} / (\underline{h} - 0.5), & \text{if } \underline{h} \geq 2. \end{cases} \tag{11}$$

As \underline{h} increases, Δy decreases, thereby indicating the presence of more refined lattice grids.

There may exist a mismatch between $\sigma(y)/\sigma^s$ and $h(y)$ in (9). This time, we measure their difference by their ratio and see how far it is away from 1. The ratio γ is defined as follows:

$$\gamma(y) \equiv \frac{\sigma(y)/\sigma^s}{h(y)} = \frac{x(y)}{h(y)}. \tag{12}$$

It is clear when h is high that γ is close to one. Therefore, the range of $\gamma(y)$ is determined by the lower bound of $h(y)$, which is \underline{h} . It can be shown that the following relations hold:

$$1 - 0.5/\underline{h} \leq \gamma(y) \leq 1 + 0.5\underline{h}, \quad \forall y, \quad \text{if } \underline{h} \geq 2, \tag{13a}$$

$$1 \leq \gamma(y) \leq 1 + 0.5\underline{h}, \quad \forall y, \quad \text{if } \underline{h} = 1, \tag{13b}$$

and

$$\max(\underline{h} - 0.5, 1) \leq x(y) \leq \underline{h} + 0.5, \quad \forall y. \tag{14}$$

Given any arbitrary lattice node y , use (5) and (9) to determine the value of k and h , respectively. One can then solve the branching probabilities p_u , p_m , and p_d at y , such that these three branches match the mean and variance of the price deviation. They can be expressed in terms of c , $\epsilon(y)$, $\gamma(y)$, and $h(y)$:

$$p_u = \frac{1}{2} \left(\frac{\epsilon^2}{h^2} + \frac{\epsilon}{h} + \frac{\gamma^2}{c^2} \right), \tag{15a}$$

$$p_d = \frac{1}{2} \left(\frac{\epsilon^2}{h^2} - \frac{\epsilon}{h} + \frac{\gamma^2}{c^2} \right), \tag{15b}$$

$$p_m = 1 - \left(\frac{\epsilon^2}{h^2} + \frac{\gamma^2}{c^2} \right). \tag{15c}$$

Proposition 1. (One-Factor Lattice Feasibility) Given c and \underline{h} , for the branching probabilities (15a)–(15c) to be legitimate for all $|\epsilon| \leq 0.5$ and γ satisfying (13a)–(13b) for all y , the following condition must hold:

$$\sqrt{\frac{\underline{h} + 0.5}{\underline{h} - 0.5}} \leq c \leq \max(\sqrt{3}, \sqrt{2\underline{h} - 1}). \tag{16}$$

Proposition 1 states how the values of the two key parameters c and \underline{h} should coordinate to achieve lattice feasibility. This proposition states that, if a lattice is configured such that its values for c and \underline{h} meet (16), the lattice will not have negative probabilities in any branches using (15a)–(15c). When $\underline{h} = 1$, the only feasible c is $\sqrt{3}$, which is an interesting number for c . When $c = \sqrt{3}$, each typical trinomial branch approximates the underlying normal distribution well (matching up to the 5th moment when $\epsilon = 0$). However, when $\underline{h} \geq 2$, c becomes flexible and its feasible value spreads over a bigger range.

2.1. Effects of Grid Refinement

As pointed in the previous section, introducing \underline{h} has an effect of grid refinement. Basically, if $\sigma(y)$ has a smaller σ^{\min} , a higher \underline{h} will be required, which leads to smaller σ^s and Δy . Since $|\mu(y)\Delta t - k(y)\Delta y| = |\epsilon(y)|\Delta y \leq 0.5\Delta y$, a smaller Δy means that the mean of the price change is better approximated. Traditionally, a discrete lattice approximates the underlying continuous processes better and better as the time step Δt decreases, which may be viewed as lattice refinement on the time space. By increasing the value of \underline{h} , we introduce a refinement on the state space of the underlying variable, even when Δt is not especially small. Therefore, grid refinement has an effect on better convergence.

2.2. Weak Convergence of the One-Factor Lattice

In the proposed trinomial lattice, the jump size $h(y)$ varies from node to node due to the state-dependent volatility. As a result, the branches may not recombine in an easily predictable way as in the constant volatility case. Therefore, it is not immediately clear that such a lattice would converge to the underlying process. Next, we show that our proposed lattice indeed converges weakly to the underlying diffusion process in (1).

Proposition 2. Let $Y_t^{\Delta t}$ denote the trinomial lattice of the one-factor process Y_t defined in (1) with $\Delta y = c\sigma^s\sqrt{\Delta t}$, and $k(y)$ and $h(y)$ defined by (5) and (9), respectively. Suppose that $Y_0^{\Delta t} = Y_0$ and the following conditions

$$0 \leq p_u, p_m, p_d \leq 1 \tag{17a}$$

$$E[(Y_{t+\Delta t}^{\Delta t} - y) | Y_t^{\Delta t} = y] = \mu(y)\Delta t \tag{17b}$$

$$E[(Y_{t+\Delta t}^{\Delta t} - y)^2 | Y_t^{\Delta t} = y] = \sigma^2(y)\Delta t + \mu^2(y)\Delta t^2, \tag{17c}$$

hold for all lattice node y and for all time interval Δt . As $\Delta t \rightarrow 0$, $Y_t^{\Delta t}$ converges weakly to the diffusion process Y_t .

2.3. Estimating σ^{\min} for CIR Model and Feller Condition

In this section, we use the Cox et al. (1985) model as an example to illustrate the applicability of the proposed framework. Consider

$$dY_t = -\lambda(Y_t - m)dt + \xi\sqrt{Y_t}dW_t, \tag{18}$$

where λ , m , and ξ are positive constants. The CIR model (18) does not meet (1) and (2), but (3a). This is because the lower bound of the square root volatility function is zero. To meet (3b), assume that $\sigma^{\min} = \varepsilon > 0$ is infinitesimal. In this section, we show how to find an approximate lower bound $\hat{\sigma}^{\min} \geq \varepsilon$ for the volatility function. Equivalently, we would find a y^{\min} such that $\hat{\sigma}^{\min} = \sigma(y^{\min}) = \xi\sqrt{y^{\min}} \geq \varepsilon$, i.e., $y^{\min} \geq \varepsilon^2/\xi^2$.

To do this, we must exploit the mean reverting (MR) property of the drift function of (18). Our approach is to find a y^{\min} , hopefully much greater than ε^2/ξ^2 , such that y^{\min} is at a level where Y_t would revert upward so that the whole lattice Y_{t+1} remains to be capped from below by it. Since the lattice is capped from below by y^{\min} , no other volatility value in the lattice is smaller than $\hat{\sigma}^{\min}$.

Note that this reverting level y^{\min} depends on Δt and decreases to 0 as $\Delta t \rightarrow 0$. This should not be a problem because, in practice, lattices yield satisfactory valuation results at some finite Δt where $\hat{\sigma}^{\min}$ hopefully is still much greater than ε .

As mentioned above, we would like to see the entire lattice to be well contained in the \mathbb{R}^+ domain. With $\hat{\sigma}^{\min} = \xi\sqrt{y^{\min}}$ and using (11), we have

$$\Delta y = \xi c\sqrt{\Delta t}\sqrt{y^{\min}} / \max(\underline{h} - 0.5, 1). \tag{19}$$

We want to find a $y^{\min} > 0$ such that

$$\left\lfloor \frac{y}{\Delta y} \right\rfloor + k(y) - h(y) \geq \left\lfloor \frac{y^{\min}}{\Delta y} \right\rfloor, \forall y \geq y^{\min} \tag{20}$$

holds. In (20), $\lfloor y/\Delta y \rfloor$ may be viewed as the index of the lattice node of y (with the index equal to 0 at $y = 0$), and from this node, $(k(y) - h(y))\Delta y$ corresponds to the price change of the downward branch. Therefore, the left-hand side of (20) represents the index of this lattice node mapped onto from y at the next time period, which should be bounded below by the node corresponding to y^{\min} . Note in (20) that y^{\min} is a continuous variable,

which may not match the value of some discrete lattice node. If $y^{\min} \geq \Delta y$, we are done; otherwise, we check if

$$k(\Delta y) - h(\Delta y) \geq 0 \tag{21}$$

in order to ensure that the lattice will revert upward at $y = \Delta y$.

An important fact is that a $y^{\min} > 0$ may not necessarily exist for any arbitrary process parameters. For example, it has been shown that $2\lambda m \geq \zeta^2$ must hold for y in (18) to be bounded below by zero, which is known as Feller condition (Feller 1951). Next, we shall show that, if the Feller condition is met, then $y^{\min} > 0$ exists and the entire lattice can be well contained in \mathbb{R}^+ .

Proposition 3. *If (i) $1 - \lambda\Delta t > 0$, and (ii) $4\lambda m > \zeta^2 c^2 / (1 - \lambda\Delta t)$ both hold, then $y^{\min} > 0$ exists and the entire lattice can be well contained in \mathbb{R}^+ .*

It is clear that condition (i) of Proposition 3 can be easily satisfied by reducing Δt , but condition (ii) may not. Compared with the Feller condition $\lambda m \geq 0.5\zeta^2$, condition (ii) has a similar form and may be viewed as the discrete version of Feller condition in our proposed lattice approach.

For a finite $\Delta t > 0$, if condition (ii) is violated, to rectify it, it is probably more effective to reduce the value of c (to its lower bound by increasing \underline{h}) than to reduce Δt . This demonstrates another application of grid refinement. As a limit, our approach requires $\lambda m > 0.25\zeta^2$ to hold (by setting $\Delta t = 0$ and $c = 1$) so that $y^{\min} > 0$ exists. This condition is slightly looser than the Feller condition due to the fact that it is easier to bound the process of y from below in the (approximate) discrete domain than in the continuous one. This also means that, whenever the Feller condition is met, a Cox et al. (1985) process can always be approximated by the proposed lattice.

Proposition 4. *Given a Cox et al. (1985) process (18) that satisfies the Feller condition, there exists a feasible lattice configuration $(\Delta t, \underline{h}, c)$ such that $y^{\min} > 0$ and the lattice can be well contained in \mathbb{R}^+ .*

3. Two-Factor Trinomial Lattice

In this section, we consider a more general two-factor model:

$$dY_{1,t} = \mu_1(Y_{1,t}, Y_{2,t})dt + \sigma_1(Y_{1,t}, Y_{2,t})dW_{1,t} \tag{22a}$$

$$dY_{2,t} = \mu_2(Y_{1,t}, Y_{2,t})dt + \sigma_2(Y_{1,t}, Y_{2,t})dW_{2,t}, \tag{22b}$$

where $W_{1,t}$ and $W_{2,t}$ are Wiener processes with an instantaneous correlation ρ . Following the treatment described in (3a)–(3b) in Section 2, we assume that there exist $\sigma_l^{\min} > 0$, $l = 1, 2$ such that

$$\sigma_l(y_1, y_2) \geq \sigma_l^{\min} > 0, \quad \forall y_1, y_2, \quad l = 1, 2. \tag{23}$$

When both individual one-factor lattices are integrated, we shall apply the same convention to all relevant notations, such as σ^s , k , h , \underline{h} , and c , by adding a subscript $l = 1, 2$ to reflect the index of the process that each notation represents.

Our idea of using two-factor trinomial lattices to approximate (22a)–(22b) follows the one proposed by Hull and White (1994): obtain a one-factor trinomial lattice for each state variable first, then integrate both lattices to one such that nine branches (3×3) emanating from each lattice node. Let $\Delta y_l = c_l \sigma_l^s \sqrt{\Delta t}$, $l = 1, 2$, a definition extended directly from that in the one-factor case in Section 2. Assume the one-factor discrete price increment, branching factor for y_l , and branch jump size are Δy_l , k_l , and h_l , $l = 1, 2$, respectively, as defined in Section 2. An example is given in Figure 1, where a node $(y_{1,t}, y_{2,t})$ at time t is shown in the left panel. The first one calculates k_l and h_l using (5) and (9), respectively, for each factor l to determine the three nodes $y_{l,t+1}^u$, $y_{l,t+1}^m$, and $y_{l,t+1}^d$ at time $t + 1$ that each factor l will map into. The nine nodes of their combinations at $t + 1$ are shown in the right panel.

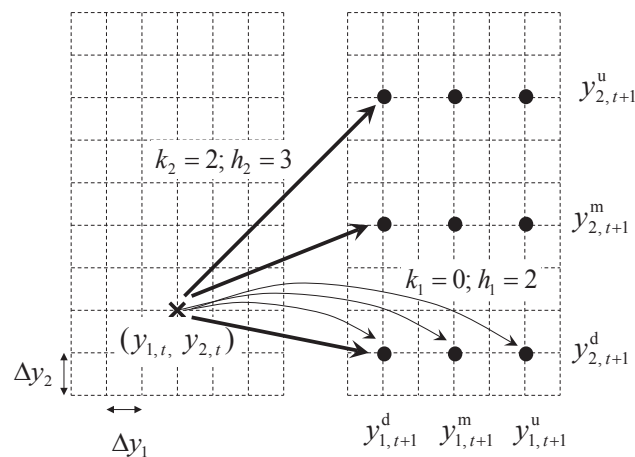


Figure 1. An example of node branching of our proposed two-factor lattice.

The main task is to solve the nine branching probabilities while matching the first two moments of each factor and their correlation. We further define $k_{lu} \equiv k_l + h_l$, $k_{lm} \equiv k_l$, and $k_{ld} \equiv k_l - h_l$, $l = 1, 2$. At each node, to determine the nine branching probabilities $p_{ij}, i, j \in \Omega$, where $\Omega \equiv \{u, m, d\}$, one solves the following linear system, denoted by $(Q_0): (Q_0)$

$$\sum_{i,j \in \Omega} p_{ij} k_{1i} \Delta y_1 = \mu_1 \Delta t, \tag{24a}$$

$$\sum_{i,j \in \Omega} p_{ij} k_{2j} \Delta y_2 = \mu_2 \Delta t, \tag{24b}$$

$$\sum_{i,j \in \Omega} p_{ij} k_{1i}^2 \Delta y_1^2 = \mu_1^2 \Delta t^2 + \sigma_1^2 \Delta t, \tag{24c}$$

$$\sum_{i,j \in \Omega} p_{ij} k_{2j}^2 \Delta y_2^2 = \mu_2^2 \Delta t^2 + \sigma_2^2 \Delta t, \tag{24d}$$

$$\sum_{i,j \in \Omega} p_{ij} k_{1i} k_{2j} \Delta y_1 \Delta y_2 = \rho \sigma_1 \sigma_2 \Delta t + \mu_1 \mu_2 \Delta t^2, \tag{24e}$$

$$\sum_{i,j \in \Omega} p_{ij} = 1, \tag{24f}$$

$$p_{ij} \geq 0, \forall i, j \in \Omega \tag{24g}$$

At each node of the trinomial lattice, it is required to solve (Q_0) to determine the branching probabilities, which has nine variables, six linear equations, and nine nonnegativity inequalities. In optimization terminology, (Q_0) is said to be *feasible* if a (feasible) solution exists that meets all constraints. Otherwise, (Q_0) is said to be *infeasible*, which means that some probabilities must be negative or greater than 1. Lattice feasibility refers to the condition where (Q_0) is feasible for *all* possible nodes in a lattice. In the next proposition, we will show that lattice feasibility is a necessary condition for weak convergence of the proposed two-factor lattice.

Proposition 5. Let $(Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t})$ denote the trinomial lattice of the two-factor process $(Y_{1,t}, Y_{2,t})$ defined in (22a)–(22b) with $\Delta y_l = c_l \sigma_l^s \sqrt{\Delta t}$, and $k_l(y)$ and $h_l(y)$ defined by (5) and (9), respectively, for $l = 1, 2$. Suppose that $(Y_{1,0}^{\Delta t}, Y_{2,0}^{\Delta t}) = (Y_{1,0}, Y_{2,0})$ and the following conditions

$$0 \leq p_{ij} \leq 1, i, j \in \Omega \tag{25a}$$

$$E[(Y_{l,t+\Delta t}^{\Delta t} - y_l) | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)] = \mu_l \Delta t, l = 1, 2 \tag{25b}$$

$$E[(Y_{l,t+\Delta t}^{\Delta t} - y_l)^2 | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)] = \sigma_l^2 \Delta t + \mu_l^2 \Delta t^2, l = 1, 2 \tag{25c}$$

$$E[(Y_{1,t+\Delta t}^{\Delta t} - y_1)(Y_{2,t+\Delta t}^{\Delta t} - y_2) | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)] = \rho \sigma_1 \sigma_2 \Delta t + \mu_1 \mu_2 \Delta t^2 \tag{25d}$$

hold for all lattice node (y_1, y_2) and for all time interval Δt . As $\Delta t \rightarrow 0$, $(Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t})$ converges weakly to the diffusion process $(Y_{1,t}, Y_{2,t})$.

3.1. Feasibility of the General Lattice

Consider the nine branches that emanated from a fixed node, where the one-factor branching probabilities are assumed to be $\{\tilde{p}_{lu}, \tilde{p}_{lm}, \tilde{p}_{ld}\}, l = 1, 2$. These probabilities are denoted with a tilde to indicate that they are not variables. In addition, assume at this node that the corresponding error factors of the branching and the jump sizes are ϵ_l and $\gamma_l, l = 1, 2$. The branching probabilities can be obtained as follows (cf. (15a)–(15c)):

$$\tilde{p}_{lu} = \frac{1}{2} \left(\frac{\epsilon_l^2}{h_l^2} + \frac{\epsilon_l}{h_l} + \frac{\gamma_l^2}{c_l^2} \right) \tag{26a}$$

$$\tilde{p}_{ld} = \frac{1}{2} \left(\frac{\epsilon_l^2}{h_l^2} - \frac{\epsilon_l}{h_l} + \frac{\gamma_l^2}{c_l^2} \right) \tag{26b}$$

$$\tilde{p}_{lm} = 1 - \left(\frac{\epsilon_l^2}{h_l^2} + \frac{\gamma_l^2}{c_l^2} \right) \tag{26c}$$

The key step is to rewrite (Q_0) to the following equivalent form (Q) , in terms of $\{\tilde{p}_{li}\}, i \in \Omega$, and $\epsilon_l, \gamma_l, l = 1, 2$: $(Q(\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2))$

$$\sum_{j \in \Omega} p_{uj} = \tilde{p}_{1u} \tag{27a}$$

$$\sum_{j \in \Omega} p_{dj} = \tilde{p}_{1d} \tag{27b}$$

$$\sum_{i \in \Omega} p_{iu} = \tilde{p}_{2u} \tag{27c}$$

$$\sum_{i \in \Omega} p_{id} = \tilde{p}_{2d} \tag{27d}$$

$$p_{uu} + p_{dd} - p_{ud} - p_{du} = \rho \frac{\gamma_1 \gamma_2}{c_1 c_2} + \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \tag{27e}$$

$$\sum_{i, j \in \Omega} p_{ij} = 1 \tag{27f}$$

$$p_{ij} \geq 0, \forall i, j \in \Omega. \tag{27g}$$

To maintain legibility, we denote $\theta_l \equiv (\epsilon_l, x_l, \gamma_l, h_l) \in \Theta_l(\underline{h}_l), l = 1, 2$, where

$$\Theta_l(\underline{h}_l) \equiv \{\theta_l \mid (7), (14), (13a) - (13b), \forall h_l \geq \underline{h}_l\}, l = 1, 2. \tag{28}$$

In (28), $\Theta_l(\underline{h}_l)$ covers all possible nodes in $Y_l, l = 1, 2$. (Q) is denoted with (θ_1, θ_2) to emphasize that it is a ‘local’ problem associated with some specific lattice node (as \tilde{p}_{lu} and \tilde{p}_{ld} are functions of $\theta_l, l = 1, 2$). Equations (27a)–(27d) intend to match the means and variances of the price deviations of the two factors, which have been met by the marginal probabilities $\{\tilde{p}_{1j}\}$ and $\{\tilde{p}_{2j}\}, j \in \Omega$. Equation (27e) is derived from (24e) with some algebra.

Consider an initial solution $p_{ij} = \tilde{p}_{1i}\tilde{p}_{2j}, i, j \in \Omega$, which satisfies (27a)–(27f) but (27e), unless $\rho = 0$. To determine the range of ρ on the right-hand side (RHS) of (27e) that (Q) is feasible, consider the following two linear programs:

$$R^{\min}(\theta_1, \theta_2) \equiv \min_{p_{ij}, i, j \in \Omega} \{p_{uu} + p_{dd} - p_{ud} - p_{du} \mid (Q) \text{ excluding (27e)}\} \tag{29a}$$

$$R^{\max}(\theta_1, \theta_2) \equiv \max_{p_{ij}, i, j \in \Omega} \{p_{uu} + p_{dd} - p_{ud} - p_{du} \mid (Q) \text{ excluding (27e)}\} \tag{29b}$$

For a given set of (θ_1, θ_2) , it is clear that (Q) is feasible, if, and only if, the RHS of (27e) is between R^{\min} and R^{\max} , i.e.,

$$R^{\min}(\theta_1, \theta_2) \leq \rho \frac{\gamma_1 \gamma_2}{c_1 c_2} + \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \leq R^{\max}(\theta_1, \theta_2). \tag{30}$$

Therefore, for $\theta_l \in \Theta_l(\underline{h}_l), l = 1, 2$, we have

$$c_1 c_2 \cdot \max_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(R^{\min}(\theta_1, \theta_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \leq \rho \leq c_1 c_2 \cdot \min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(R^{\max}(\theta_1, \theta_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right), \tag{31}$$

Given the values of c_1 and c_2 , using (31), the range of the correlation ρ between the two factors for which the lattice can guarantee feasible branching probabilities at all nodes and all stages can be identified. Next, we will show that (31) is symmetric such that its upper bound is the negative of its lower bound.

Proposition 6. For any arbitrary $\theta_l = (\epsilon_l, x_l, \gamma_l, h_l) \in \Theta_l(\underline{h}_l), l = 1, 2$, the following equality holds:

$$\min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(R^{\max}(\theta_1, \theta_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) = - \max_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(R^{\min}(\theta_1, \theta_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right), l = 1, 2. \tag{32}$$

Since (31) is symmetric, we can focus on solving R^{\max} . The closed-form expression for R^{\max} has been derived in Tseng and Lin (2007), which is duplicated in the following proposition for the sake of clarity.

Proposition 7. $R^{\max} = \min\{R_1^{\max}, R_2^{\max}, R_3^{\max}, R_4^{\max}, R_5^{\max}, R_6^{\max}\}$, where

$$\begin{aligned} R_1^{\max} &= \tilde{p}_{1u} + \tilde{p}_{1d} \\ R_2^{\max} &= \tilde{p}_{2u} + \tilde{p}_{2d} \\ R_3^{\max} &= \tilde{p}_{1u} + \tilde{p}_{2d} \\ R_4^{\max} &= 1 - (\tilde{p}_{1d} - \tilde{p}_{2d}) - (\tilde{p}_{2u} - \tilde{p}_{1u}) \\ R_5^{\max} &= \tilde{p}_{2u} + \tilde{p}_{1d} \\ R_6^{\max} &= 1 - (\tilde{p}_{2d} - \tilde{p}_{1d}) - (\tilde{p}_{1u} - \tilde{p}_{2u}), \end{aligned}$$

where $\tilde{p}_{li}, l = 1, 2, i \in \Omega$ are from (26a)–(26c).

Note that $R_l^{\max}, l = 1, \dots, 6$ are also functions of θ_1 and θ_2 . It can be seen that R_1^{\max}, R_3^{\max} , and R_4^{\max} reduce to R_2^{\max}, R_5^{\max} , and R_6^{\max} , and vice versa, respectively, by exchanging the factor indices 1 and 2.

Using the result from Proposition 6 to find the upper bound of (31), one needs to solve

$$\min_{\theta_1, \theta_2} \min_{i=1, \dots, 6} \frac{1}{\gamma_1 \gamma_2} \left\{ R_i^{\max} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right\} \tag{33}$$

for $\theta_l \in \Theta_l(\underline{h}_l), l = 1, 2$. It would be easier to solve if one could switch the two minimization operators in (33) as follows:

$$\min_{i=1, \dots, 6} \left\{ \min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} (R_i^{\max} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2}) \right\}. \tag{34}$$

It turns out that both (33) and (34) are equivalent, which is shown in the following proposition.

Proposition 8. Let $f_1, \dots, f_n : \mathbb{R}^n \rightarrow \mathbb{R}$ be n continuous functions and $D \subseteq \mathbb{R}^n$, where $n \in \mathbb{N}$ is finite. Then,

$$\min_{z \in D} \min\{f_1(z), \dots, f_n(z)\} = \min\{\min_{z \in D} f_1(z), \dots, \min_{z \in D} f_n(z)\}. \tag{35}$$

Next, we state the main theorem of this section.

Theorem 1. (Two-Factor Lattice Feasibility) Given a lattice configuration $(c_1, c_2, \underline{h}_1, \underline{h}_2)$, (Q) is feasible for all $\theta_l \in \Theta_l(\underline{h}_l), l = 1, 2$, if, and only if, $|\rho| \leq \bar{\rho}(c_1, c_2, \underline{h}_1, \underline{h}_2)$, where

$$\bar{\rho}(c_1, c_2, \underline{h}_1, \underline{h}_2) \equiv c_1 c_2 \cdot \min\{w_1, w_2, w_3, w_4\}, \tag{36}$$

and

$$w_1 = \min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(\frac{\phi_1^2}{h_1^2} - \frac{\phi_1}{2h_1 h_2} + \frac{\gamma_1^2}{c_1^2} \right), \quad \phi_1(h_1, h_2) \equiv \min\left(\frac{h_1}{4h_2}, \frac{1}{2}\right), \tag{37a}$$

$$w_2 = \min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(\frac{\phi_2^2}{h_2^2} - \frac{\phi_2}{2h_1 h_2} + \frac{\gamma_2^2}{c_2^2} \right), \quad \phi_2(h_1, h_2) \equiv \min\left(\frac{h_2}{4h_1}, \frac{1}{2}\right), \tag{37b}$$

$$w_3 = \min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left[\frac{1}{8}(\phi_3^2 - 1) + \frac{1}{2} \left(\frac{\gamma_1^2}{c_1^2} + \frac{\gamma_2^2}{c_2^2} \right) \right],$$

$$\phi_3(h_1, h_2) \equiv \max\left(1 - \frac{1}{h_1} - \frac{1}{h_2}, 0\right), \tag{37c}$$

$$w_4 = \min_{\theta_1, \theta_2} \frac{1}{\gamma_1 \gamma_2} \left(1 - \frac{1}{2h_1}\right) \left(1 - \frac{1}{2h_2}\right). \tag{37d}$$

Theorem 1 is general, but, unfortunately, no explicit functional expressions for w_1, \dots, w_4 are available because each of them requires solving a two-dimensional global minimum of a discontinuous function. However, given a set of lattice parameters $(c_1, c_2, \underline{h}_1, \underline{h}_2)$, using numerical methods, such as exhaustive search, one can easily obtain the numerical values of w_1, \dots, w_4 .

From the perspective of minimizing computational requirements, one would prefer smaller values of $(\underline{h}_1, \underline{h}_2)$ so that the grid size may not become too small. To achieve that, next, we try to maximize $\bar{\rho}$ over c_1 and c_2 for a given $(\underline{h}_1, \underline{h}_2)$ pair:

$$\rho_{\max}(\underline{h}_1, \underline{h}_2) = \max_{c_1, c_2} \{\bar{\rho}(c_1, c_2, \underline{h}_1, \underline{h}_2) | c_1, c_2 \text{ subject to (16)}\}. \tag{38}$$

Using numerical methods, Table 1 shows the values of ρ_{\max} for all 100 pairs of $(\underline{h}_1, \underline{h}_2)$ for $1 \leq \underline{h}_1, \underline{h}_2 \leq 10$. Note that, since ρ_{\max} is symmetric, Table 1 only displays half of the pairs with $\underline{h}_1 \geq \underline{h}_2$. From Table 1, it is clear that $\rho_{\max}(\underline{h}_1, \underline{h}_2)$ is an increasing function in \underline{h}_1 and \underline{h}_2 . If one considers increasing \underline{h}_1 by 1 to be as difficult in terms of computational effort as increasing \underline{h}_2 by 1, then the diagonal elements where $(\underline{h}_1 = \underline{h}_2)$ seem to be the most efficient choices.

Table 1. The 100 values of $\rho_{\max}(h_1, h_2)$ for $1 \leq h_1, h_2 \leq 10$.

(h_1, h_2)	1	2	3	4	5	6	7	8	9	10
1	0.3333	0.4957	0.5544	0.5866	0.6076	0.6222	0.6319	0.6343	0.6362	0.6376
2		0.6399	0.6892	0.7156	0.7327	0.7450	0.7539	0.7610	0.7668	0.7714
3			0.7401	0.7661	0.7834	0.7947	0.8050	0.8112	0.8170	0.8212
4				0.7949	0.8124	0.8235	0.8321	0.8393	0.8450	0.8499
5					0.8302	0.8414	0.8504	0.8571	0.8633	0.8682
6						0.8551	0.8628	0.8707	0.8751	0.8798
7							0.8735	0.8797	0.8856	0.8898
8								0.8878	0.8919	0.8961
9									0.8991	0.9033
10										0.9084

When $h_1 = h_2$, to obtain the optimal solutions (c_1^*, c_2^*) for achieving ρ_{\max} in (38), it turns out that $w_1 = w_2$ in Theorem 1 are the smallest elements in the minimum operator in (36), so a symmetrical optimal solution $c_1^* = c_2^*$ is obtained. Using numerical methods, the values of ρ_{\max} and the corresponding (c_1^*, c_2^*) are summarized in Table 2.

Table 2. The values of ρ_{\max} and the optimal (c_1^*, c_2^*) when $h_1 = h_2$.

$h_1 = h_2$	$\rho_{\max}(h_1, h_2)$	(c_1^*, c_2^*)
1	0.3333	(1.7321, 1.7321)
2	0.6397	(1.3340, 1.3340)
3	0.7400	(1.2052, 1.2052)
4	0.7949	(1.1469, 1.1469)
5	0.8320	(1.1145, 1.1145)
6	0.8551	(1.0931, 1.0931)
7	0.8735	(1.0792, 1.0792)
8	0.8878	(1.0686, 1.0686)
9	0.8991	(1.0602, 1.0602)
10	0.9084	(1.0543, 1.0543)

As an example, if $\rho = -0.8$, using the optimized (c_1^*, c_2^*) from Table 2, one can use $h_1 = h_2 = 5$ by using $(c_1, c_2) = (1.1145, 1.1145)$ obtained from the Appendix A. Note that the results presented in this section are for general two-factor lattices. If the two underlying diffusion processes have some special structures, e.g., the class of square root volatility models, then ρ_{\max} may be further increased so that the required (h_1, h_2) that guarantees lattice feasibility may be lowered. In the next section, we focus on the Heston SV model, for which explicit functional expressions for w_1, \dots, w_4 are available.

3.2. Lattice for the Heston SV Model

Consider the Heston SV model (Heston 1993) as follows:

$$dY_{1,t} = \mu dt + \sqrt{Y_{2,t}} dW_{1,t} \tag{39a}$$

$$dY_{2,t} = -\lambda(Y_{2,t} - m)dt + \xi \sqrt{Y_{2,t}} dW_{2,t}, \tag{39b}$$

where $Y_{1,t} = \ln S_t$ is the logarithm of the stock price; μ, λ, m , and ξ are positive constants.

Since both volatilities in (39a) and (39b) are functions of Y_2 , we focus on the MR process of Y_2 . As mentioned in Section 2.2, if the Feller condition is satisfied, $y_2^{\min} > 0$ exists for (39b). Based on y_2^{\min} , we can identify positive σ_2^{\min} and σ_2^s . Once σ_2^s is obtained, we set

$$\sigma_1^s = \sigma_2^s / \xi. \tag{40}$$

Proposition 9. To implement the Heston SV model given in (39a)–(39b) using the proposed trinomial lattice and (40), if $h_1 = h_2$, then $h_1 = h_2$, $x_1 = x_2$, and $\gamma_1 = \gamma_2$ at all nodes.

Using Proposition 9, we will show that Theorem 1 has an explicit functional form for $\bar{\rho}$ if we set $\underline{h}_1 = \underline{h}_2$. Since μ_1 is a constant, ϵ_1 is a fixed constant for all nodes. We denote it as $\tilde{\epsilon}_1$. Without loss of generality, we assume $\tilde{\epsilon}_1 = 0$. To make $\tilde{\epsilon}_1 = 0$, from (5), we need to have

$$\frac{\mu_1 \Delta t}{c_1 \sigma_1^s \sqrt{\Delta t}} = \frac{\mu_1 \sqrt{\Delta t}}{c_1 \sigma_1^s} = \text{the nearest integer} = k_1. \tag{41}$$

This can be achieved by adjusting the value of c_1 , σ_1^{\min} , or Δt . Adjusting c_1 and/or Δt is more straightforward than adjusting σ_1^{\min} . However, since c_1 is a parameter for lattice configuration, we recommend adjusting the value of Δt slightly to eliminate the remainder of (41) in order to make $\tilde{\epsilon}_1 = 0$.

With $\epsilon_1 = 0$ and $\underline{h}_1 = \underline{h}_2 = \underline{h}$, using (31), the condition for lattice feasibility can be significantly simplified as follows:

$$\max_{\theta_1, \theta_2} \frac{1}{\gamma^2} \left(R^{\min}(\theta_1, \theta_2) \right) \leq \frac{\rho}{c_1 c_2} \leq \min_{\theta_1, \theta_2} \frac{1}{\gamma^2} \left(R^{\max}(\theta_1, \theta_2) \right), \tag{42}$$

for $\theta_l \in \Theta_l(\underline{h}), l = 1, 2$.

Theorem 2. (Lattice feasibility for the Heston SV) Pertaining to the Heston SV model in (39a)–(39b), assume $\underline{h}_1 = \underline{h}_2 = \underline{h}$ and $\epsilon_1 = 0$. Given a lattice configuration $(c_1, c_2, \underline{h})$, (Q) is feasible for all $\theta_l \in \Theta_l(\underline{h}), l = 1, 2$, if, and only if, $\rho \leq \bar{\rho}^H(c_1, c_2, \underline{h})$, where

$$\bar{\rho}^H(c_1, c_2, \underline{h}) = c_1 c_2 \cdot \min\{w_1^H, w_2^H, w_3^H, w_4^H\}, \tag{43}$$

where

$$w_1^H = \frac{1}{c_1^2} \tag{44a}$$

$$w_2^H = \frac{1}{c_2^2} \tag{44b}$$

$$w_3^H = \frac{1}{2} \left(\frac{1}{c_1^2} + \frac{1}{c_2^2} \right) - \frac{1}{4(\max(\underline{h}, 2) - 0.5)}, \tag{44c}$$

$$w_4^H = \frac{\underline{h}(\underline{h} - 0.5)}{(\underline{h} + 0.5)^2} \tag{44d}$$

Next, we try to maximize $\bar{\rho}^H$ by adjusting the lattice configuration (c_1, c_2) while maintaining lattice feasibility. Let

$$\rho_{\max}^H(\underline{h}) = \max_{c_1, c_2} \{ \bar{\rho}^H(c_1, c_2, \underline{h}) \mid c_1, c_2 \text{ subject to (16)} \}. \tag{45}$$

Using numerical methods, the solution of (45) is obtained as follows. When $\underline{h} \leq 2$, ρ_{\max}^H is achieved at the lower bounds of c_1 and c_2 , where $c_1 = c_2 = \sqrt{(\underline{h} + 0.5)/(\underline{h} - 0.5)}$, and ρ_{\max}^H is determined by w_3^H . When $\underline{h} \geq 3$, ρ_{\max}^H is achieved at $c_1 = \sqrt{(\underline{h} + 0.5)/(\underline{h} - 0.5)}$, the lower bound, and c_2 at the point where $w_3^H = w_4^H$. That is,

$$\rho_{\max}^H(\underline{h}) = c_1^H(\underline{h}) c_2^H(\underline{h}) \frac{\underline{h}(\underline{h} - 0.5)}{(\underline{h} + 0.5)^2}, \quad \underline{h} \geq 3, \tag{46}$$

where

$$c_1^H(\underline{h}) = \sqrt{\frac{\underline{h} + 0.5}{\underline{h} - 0.5}} \tag{47}$$

$$c_2^H(\underline{h}) = \sqrt{\frac{(\underline{h} + 0.5)^2(\underline{h} - 0.5)}{\underline{h}^3 - \underline{h}^2 + 1.25\underline{h}}}. \tag{48}$$

It can be seen that $\rho_{\max}^H(\underline{h})$ approaches 1 as \underline{h} increases. The value of $\rho_{\max}^H(\underline{h})$ for $\underline{h} = 1, \dots, 10$ is given in Table 3, along with the corresponding (c_1^H, c_2^H) . Note that $\rho_{\max}^H(\underline{h})$ is symmetric in c_1 and c_2 . Therefore, another solution for $\rho_{\max}^H(\underline{h})$ is to switch c_1^H and c_2^H .

Table 3. The values of ρ_{\max}^H (with $\tilde{\epsilon}_1 = 0$).

\underline{h}	$\rho_{\max}^H(\underline{h})$	(c_1^H, c_2^H)	\underline{h}	$\rho_{\max}^H(\underline{h})$	(c_1^H, c_2^H)
1	0.5000	(1.7321, 1.7321)	6	0.9453	(1.0871, 1.1133)
2	0.7222	(1.2910, 1.2910)	7	0.9549	(1.0742, 1.0989)
3	0.8596	(1.1832, 1.1866)	8	0.9616	(1.0646, 1.0877)
4	0.9065	(1.1339, 1.1564)	9	0.9667	(1.0572, 1.0787)
5	0.9308	(1.1055, 1.1319)	10	0.9705	(1.0513, 1.0714)

Back to the previous example, if $\rho = -0.8$, now it only requires $\underline{h}_1 = \underline{h}_2 = 3$ with $(c_1, c_2) = (1.1832, 1.1866)$ or $(1.1866, 1.1832)$ to achieve lattice feasibility, a reduction from $\underline{h}_1 = \underline{h}_2 = 5$ of the optimized general model.

4. Impact of Lattice Infeasibility on Option Valuation

In this section, we investigate how much lattice infeasibility could impact option valuation. Consider the following SV model of Heston (1993) where the stock price (S_t) and the instantaneous variance (V_t) under the risk neutral measure are defined as follows:

$$d \ln S_t = (r - q - \frac{V_t}{2})dt + \sqrt{V_t}dW_{1,t}, \tag{49a}$$

$$dV_t = \lambda(m - V_t)dt + \xi\sqrt{V_t}dW_{2,t}, \tag{49b}$$

where r is the constant risk-free rate, q is the dividend yield, and $dW_{1,t}$ and $dW_{2,t}$ are Wiener processes such that $dW_{1,t} \cdot dW_{2,t} = \rho dt$.

4.1. An Optimization Perspective

The problem (Q) contains six linear equations with nine variables and a nonnegativity constraint. Since (Q) has no linear independence of the linear equations, these linear equations have infinitely many solutions (sets of branching probabilities). Therefore, the key is the nonnegativity constraint, which requires a solution to be a set of legitimate probabilities. When the term ‘lattice feasibility’ was coined by Tseng and Lin (2007), there was an implication of using optimization to determine branching probabilities. The idea was to add an objective function to be optimized subject to (Q). Since not all feasible solutions contribute the same to the objective function, using optimization would find not only a feasible solution, but an optimal solution. The objective function in this context refers to the quality of fitting the probability distribution of the underlying uncertainties. The authors in Tseng and Lin (2007) used the following objective function:

$$\min_{p_{ij}} \sum_{i,j \in \Omega} (p_{ij} - \tilde{p}_{1i}\tilde{p}_{2j})^2, \tag{50}$$

where $p_{ij}, i, j \in \Omega$ are subject to (Q); and \tilde{p}_{1i} and \tilde{p}_{2j} are marginal probabilities obtained from (26a)–(26c). Tseng and Lin (2007) showed that doing so best fits the probability distribution of the underlying variables in some measure.

The optimization problem in (50) is a standard quadratic programming (QP) problem with linear constraints. Apparently, the optimal solution of (50) is $p_{ij} = \tilde{p}_{1i}\tilde{p}_{2j}, i, j \in \Omega$ when $\rho = 0$. When $\rho \neq 0$, this optimization problem finds the solution that has the least deviation from the solution of the uncorrelated case. At each node, their approach requires solving a simple optimization problem to determine branching probabilities for (Q). We adopt this approach in this paper and refer to it as Best-Fit. Although solving a (QP) at each node may seem cumbersome, we have developed an iterative method to identify the binding constraints at optimality. Once the binding constraints are identified,

the corresponding feasible solution is the optimal one due to the convexity of the QP. This approach is very efficient as it usually takes a few trials to correctly identify the binding constraints.

On the other hand, we need a counterpart method for determining branching probabilities that works well in practice but may not guarantee lattice feasibility. We consider the popular lattice approach proposed by Hull and White (1994) (denoted by H&W) as this counterpart method. Note that the comparison is conducted in the same lattice framework such that the lattice structure is exactly the same except for their respective methods of determining branching probabilities at each node: the first is by Best-Fit and the second by H&W.

When $\rho < 0$, Hull and White (1994) suggests with $\delta = \rho/36$:

$$\begin{bmatrix} p_{du} & p_{mu} & p_{uu} \\ p_{dm} & p_{mm} & p_{um} \\ p_{dd} & p_{md} & p_{ud} \end{bmatrix} = \begin{bmatrix} \tilde{p}_{1d}\tilde{p}_{2u} & \tilde{p}_{1m}\tilde{p}_{2u} & \tilde{p}_{1u}\tilde{p}_{2u} \\ \tilde{p}_{1d}\tilde{p}_{2m} & \tilde{p}_{1m}\tilde{p}_{2m} & \tilde{p}_{1u}\tilde{p}_{2m} \\ \tilde{p}_{1d}\tilde{p}_{2d} & \tilde{p}_{1m}\tilde{p}_{2d} & \tilde{p}_{1u}\tilde{p}_{2d} \end{bmatrix} + \begin{bmatrix} -\delta & -4\delta & +5\delta \\ -4\delta & +8\delta & -4\delta \\ +5\delta & -4\delta & -\delta \end{bmatrix}; \quad (51a)$$

and similarly when $\rho > 0$,

$$\begin{bmatrix} p_{du} & p_{mu} & p_{uu} \\ p_{dm} & p_{mm} & p_{um} \\ p_{dd} & p_{md} & p_{ud} \end{bmatrix} = \begin{bmatrix} \tilde{p}_{1d}\tilde{p}_{2u} & \tilde{p}_{1m}\tilde{p}_{2u} & \tilde{p}_{1u}\tilde{p}_{2u} \\ \tilde{p}_{1d}\tilde{p}_{2m} & \tilde{p}_{1m}\tilde{p}_{2m} & \tilde{p}_{1u}\tilde{p}_{2m} \\ \tilde{p}_{1d}\tilde{p}_{2d} & \tilde{p}_{1m}\tilde{p}_{2d} & \tilde{p}_{1u}\tilde{p}_{2d} \end{bmatrix} + \begin{bmatrix} +5\delta & -4\delta & -\delta \\ -4\delta & +8\delta & -4\delta \\ -\delta & -4\delta & +5\delta \end{bmatrix}. \quad (51b)$$

4.2. Numerical Comparisons: Best-Fit vs. H&W

It is well known that the feasibility of (Q) becomes harder to meet when the value of the instantaneous correlation ρ is high. If one allows some branching probabilities to be negative, the corresponding probability distribution(s) of the price and/or the volatility is distorted. Depending on the degree of the distortion, there may be some impact on the option valuation. We use both approaches (Best-Fit and H&W) to value a European call option under the Heston SV model. The parameters are taken from Table 1 of Ball and Roma (1994) with $r = q = 0$, $\lambda = 8$, $m = V_0 = 0.1225$, and $\xi = 0.8$.

With $S_0 = \$100$, three cases, corresponding to three different strike prices, $K = \$80$, $\$100$, and $\$120$, are tested with the correlation ρ ranging between -0.8 and 0.8 . The result is summarized in Table 4. For the lattice, we consider $n = 100$ (with $T = 6$ months) to make sure that both methods (Best-Fit and H&W) can best fit the probability distribution of the underlying variables.

Table 4. Impact of lattice infeasibility on options pricing.

ρ	In-the-Money (ITM)			At-the-Money (ATM)			Out-of-the-Money (OTM)			Infeasibility (%)	
	Exact	Best-Fit	H&W	Exact	Best-Fit	H&W	Exact	Best-Fit	H&W	Best-Fit	H&W
0.8	21.5729	21.5704	21.9341	9.8203	9.8259	9.7892	4.3276	4.3337	3.9417	0	99.79
0.7	21.6637	21.6620	21.9489	9.8140	9.8194	9.7859	4.2401	4.2454	3.9227	0	1.03
0.6	21.7499	21.7490	21.9495	9.8060	9.8107	9.7854	4.1509	4.1553	3.9221	0	0.21
0.5	21.8320	21.8313	21.9926	9.7965	9.8010	9.7776	4.0600	4.0640	3.8665	0	0.21
0.4	21.9104	21.9100	22.0347	9.7853	9.7903	9.7694	3.9672	3.9710	3.8103	0	0.11
0.3	21.9854	21.9853	22.0757	9.7726	9.7779	9.7606	3.8725	3.8760	3.7534	0	0.00
0.2	22.0571	22.0575	22.1158	9.7584	9.7640	9.7514	3.7757	3.7790	3.6958	0	0.00
0.1	22.1260	22.1267	22.1549	9.7428	9.7485	9.7417	3.6769	3.6798	3.6375	0	0.00
0.0	22.1920	22.1931	22.1931	9.7256	9.7315	9.7315	3.5759	3.5784	3.5784	0	0.00
-0.1	22.2555	22.2569	22.2304	9.7070	9.7129	9.7210	3.4726	3.4747	3.5186	0	0.00
-0.2	22.3165	22.3183	22.2669	9.6870	9.6928	9.7101	3.3668	3.3684	3.4579	0	0.00
-0.3	22.3752	22.3773	22.3026	9.6655	9.6712	9.6987	3.2584	3.2596	3.3964	0	0.00
-0.4	22.4317	22.4341	22.3375	9.6425	9.6481	9.6869	3.1472	3.1480	3.3339	0	0.11
-0.5	22.4861	22.4888	22.3717	9.6181	9.6234	9.6745	3.0330	3.0334	3.2706	0	0.21
-0.6	22.5386	22.5414	22.4051	9.5922	9.5969	9.6617	2.9157	2.9156	3.2063	0	0.21
-0.7	22.5891	22.5925	22.4052	9.5647	9.5700	9.6619	2.7949	2.7940	3.2057	0	1.03
-0.8	22.6379	22.6415	22.4165	9.5357	9.5410	9.6579	2.6703	2.6687	3.1832	0	99.79

The last two columns of Table 4 show the percentage of lattice nodes (from $t = 0$ to T) that each has at least one outgoing branch with a probability either negative or exceeding unity. Note that this should not be confused with the situation where a price node at some stage has a negative probability to prevail. A lattice node may receive many incoming branches from other nodes in the previous stage. While some of the incoming branches may have negative probabilities, it is unlikely that the resultant probability for reaching this node is negative.

Basically, the option prices using Best-Fit are very close to the exact values, within one cent. It can be seen that Best-Fit maintains lattice feasibility for all ρ values tested, while H&W meets the feasibility condition only when $|\rho| \leq 0.3$. When $0.3 < |\rho| \leq 0.7$, there is only a very small number of nodes containing branches with negative probabilities. However, when $|\rho| = 0.8$, almost all the nodes (99.79%) violate the feasibility condition. This justifies our selection of H&W as the counterpart, as this method indeed provides very good approximations of branching probabilities that solve (Q). Figure 2 displays the percentage deviation of the obtained option prices from the exact ones by changing ρ using Best-Fit and H&W.

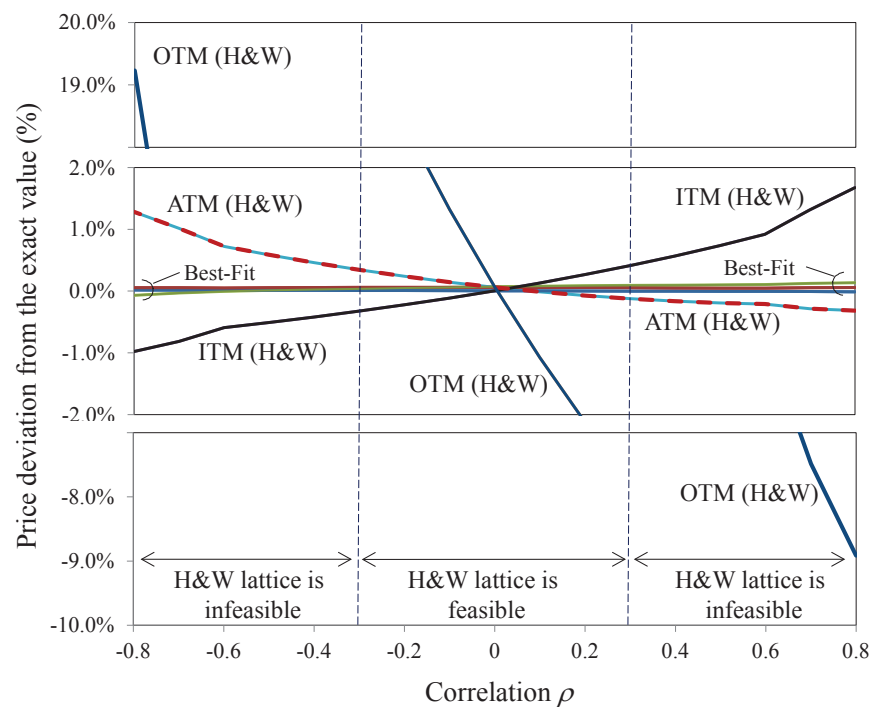


Figure 2. Deviation (%) of the values of European call options from the exact ones using Best-Fit and H&W for branching probabilities.

It can be seen from Figure 2 that H&W obtains a precise value only when there is no correlation. As $|\rho|$ increases, the option value obtained by H&W starts to deviate from the exact values. The deviation is roughly a piecewise linear function of ρ , whose slope doubles when $|\rho| > 0.6$. On the other hand, the deviations of Best-Fit are largely confined within 0.6 cents (less than 0.1%). Comparing both methods, we make the following three observations for H&W’s method:

- Consider the OTM option and the ITM option when $|\rho| \leq 0.3$. Though lattice feasibility is fully maintained, the error persists for $\rho \neq 0$. Therefore, lattice feasibility alone cannot guarantee good valuation, especially with a finite time step Δt . The valuation error would converge to zero only when Δt is sufficiently small. Therefore, lattice feasibility is merely a necessary condition for accurate valuations. From this perspective, the optimization approach for finding a feasible and optimal set of branching probabilities makes sense.

- Consider the ATM option. Its pricing errors are relatively small for all ρ values tested. Even when $\rho = |0.8|$, where infeasibility occurs at almost all nodes, the error does not seem too bad (0.3% when $\rho = 0.8$; 1.3% when $\rho = -0.8$). This shows that sometimes negative probabilities seem to matter less.
- Lattice infeasibility means that there are some distortions on the probability distribution of the underlying uncertainties. The effect can be overvaluation (e.g., OTM and ATM options when $\rho < 0$) or undervaluation (e.g., ITM option when $\rho < 0$) of the options.

In Figure 3, we plot the exact probability density functions (PDF) of the (logarithm of the) stock price (dotted curve) at different correlation levels, and the PDF approximated by the lattice using H&W (solid curve) and Best-Fit for branching probabilities. The exact PDF is obtained from a standard integration scheme of characteristic functions (e.g., see [Rough 2013](#)). The probability distributions are taken when $\rho = \pm 0.3, \pm 0.6$, and ± 0.8 . At $\rho = \pm 0.3$, both Best-Fit and H&W achieve lattice feasibility (with the first two moments of price and volatility deviations matched). There is no visible distortion in the PDF from Figure 3, yet its OTM option price still has an error of more than 3% (12 cents). This indicates that the optimization of (50) subtly improves the approximation of the tail distribution.

When $|\rho| > 0.3$, the discrepancy between the exact PDF and the PDF approximated by H&W becomes visible. Using $\rho = \pm 0.8$ in Figure 3 as an example, where the discrepancy is most distinct, the exact PDF and the one approximated by Best-Fit are still visually indistinguishable. In general, when $\rho > 0.3$ (or $\rho < -0.3$), it can be seen that, using H&W, the price distribution is distorted such that it is slightly less skewed to the right (or left) with a bigger (or smaller) mode. In Figure 3, the three strike prices tested corresponding to OTM, ATM, and ITM options are also identified. When $\rho > 0.3$ (or $\rho < -0.3$), it can be seen that using this distorted price distribution to value a European call option undervalues (or overvalues) the OTM and ATM ones, but overvalues (or undervalues) the ITM option. We make the following three additional observations:

- The price of an OTM option is directly impacted by lattice infeasibility as its value is determined by the tail distribution, which is the part of the probability distribution that is most affected by negative probabilities.
- For an ITM option, a wider part of the probability distribution becomes relevant, which tends to involve both tails and the central part, making the overall effects hard to predict.
- For an ATM option, the distorted tail part seems less important because the less distorted central part dominates the valuation.

To sum up, since, in reality, *how* the underlying distribution is distorted by lattice infeasibility is unknown a priori, it seems unlikely that one could exploit negative probabilities. However, if one really hopes for negative probabilities to work to their advantage, it is less likely to happen on OTM options, but more probable on ATM options.

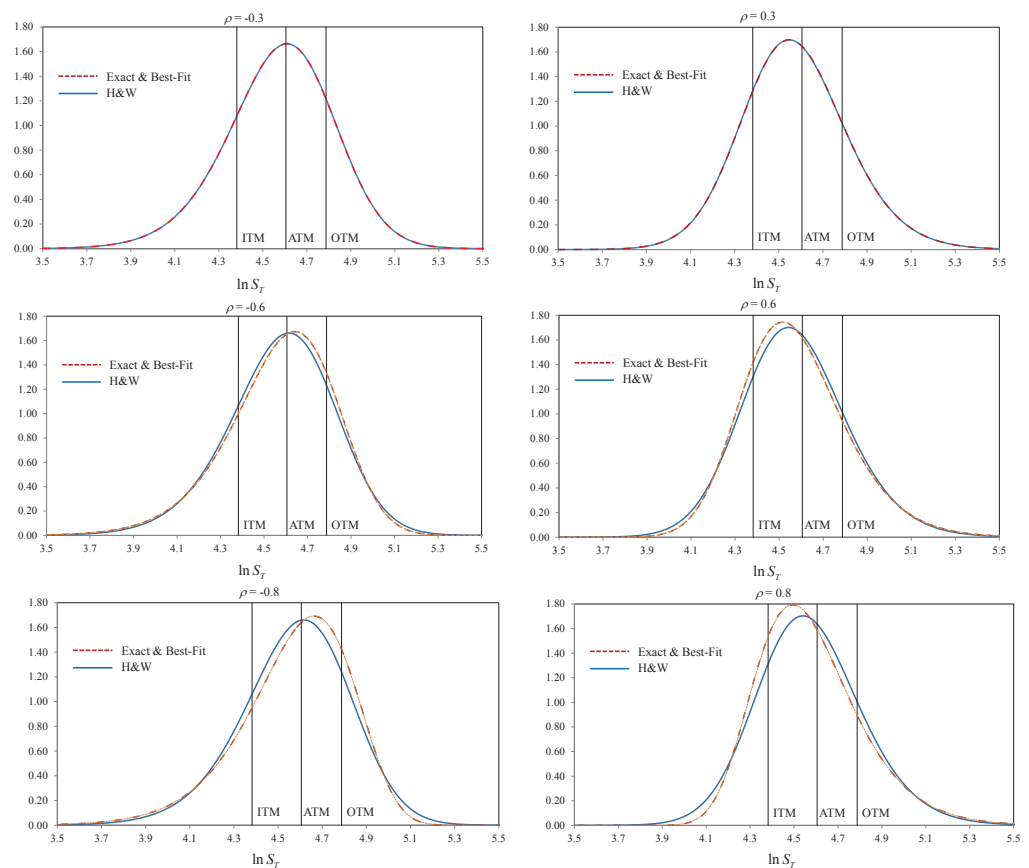


Figure 3. Illustration of the underlying distributions of $\ln S_T$.

5. Performance of the Best-Fit Lattice

In this section, we provide more numerical results for a lattice equipped with lattice feasibility and the Best-Fit method for branching probabilities. We continue to focus on the Heston SV model in (49a)–(49b). Since this approach guarantees feasible branching probabilities that also fit the underlying probability distribution well, with no surprise, all results indicate that such an approach is very reliable for accurate option valuations.

5.1. European Options Valuation

To make a comprehensive analysis of pricing errors, we compare the option prices obtained by the proposed lattice model with the exact solutions of European call options of various specifications. These specifications are drawn from combinations of the following factors: $S_0 = 100$, $q = 2\%$ (dividend yield), $T = 0.5$ years (time to maturity), $m = 0.04$, $\lambda \in \{2, 4, 8\}$, $\zeta \in \{0.3, 0.6, 0.9\}$, $K \in \{90, 100, 110\}$ (strike price), $r \in \{3\%, 5\%, 7\%\}$, $V_0 \in \{0.02, 0.04, 0.08\}$, and $\rho \in \{-0.5, -0.25, 0, 0.25, 0.5\}$. Those combinations that do not meet the Feller condition (i.e., $2\lambda m \geq \zeta^2$) are excluded. After that, 540 sets remain for the testing. The accuracy measure used in this paper is the root mean squared error (RMSE), defined as follows:

$$RMSE = \sqrt{\frac{1}{540} \sum_{i=1}^{540} e_i^2}, \tag{52}$$

where the error e_i is the difference between the exact value of the i -th European call option and the estimated option value using the proposed lattice model.

From Table 5, it can be seen that the pricing errors of the proposed lattice model are generally small compared with the exact option prices even if the number of time steps (N) is small. For instance, when $N = 10$, the RMSE of the proposed trinomial method for pricing European call options is \$0.0061, which is far smaller than the bid-ask spreads observed in the market. Moreover, Table 5 indicates that the rate of convergence of the proposed method is of order $O(1/N)$.

Table 5. The accuracy of the proposed Best-Fit model for pricing European call options under the Heston SV model.

No. of Time Steps (N)	10	20	30	40
RMSE	0.0061	0.0027	0.0024	0.0013

5.2. Convergence and Complexity

Consider another test case taken from Table 1 of Ball and Roma (1994) with $r = q = 0$, $\rho = 0$, $\lambda = 4$, $m = V_0 = 0.09$, and $\xi = 0.4$. In addition, assume $T = 0.5$ months, $S_0 = K = \$100$. Using this example, we also investigate the convergence pattern of the option prices obtained by the proposed lattice approach and the exact value (\$8.3595) when the number of time steps increases, along with its computational requirement. The results are presented in Figure 4. As seen in the upper portion of Figure 4, the option prices obtained by the proposed lattice model do converge to the exact price. We have investigated the convergence pattern for more than twenty cases, and the results are similar to those shown in Figure 4. In the lower portion, we show the CPU times required for obtaining the option values. The CPU times (in seconds) are measured on a personal computer with Intel Core 2 Duo processor E8400 of 3 GHz. We also record the number of lattice nodes in the final stage. Both the CPU times and the node numbers are displayed in logarithmic scales on both the horizontal and vertical axes. The purpose is to check whether the computational complexity is exponential. A linear behavior in a log-log graph, such as the lower portion of Figure 4, indicates that the complexity is polynomial. It is estimated for this particular instance that the CPU time is approximately of order $O(N^4)$ and the number of nodes in the final stage is of order $O(N^3)$, where N is the number of steps. The result indicates that, as N increases, most branches do recombine, which prevents the number of lattice nodes from growing exponentially with N . In this example, when $N = 30$, the option value is already within 0.2% of the exact value using about three seconds of CPU time and about 40,000 nodes in the final stage. When $N = 45$, the pricing error of the proposed approach is within 0.1%, using about 21 s of CPU time and 160,000 lattice nodes in the final stage. The number of lattice nodes involved in the proposed approach is indeed much greater than that of using traditional recombining lattices. However, the computation using the proposed lattice model can be managed to be quite efficient.

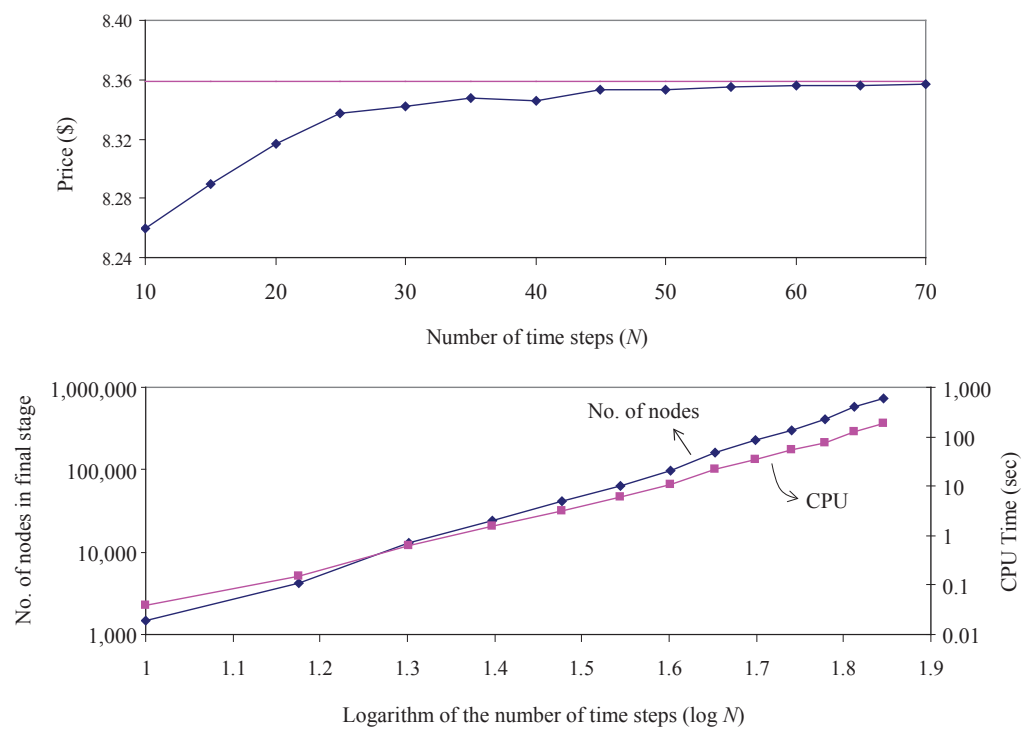


Figure 4. Convergence pattern of the option prices and the computational requirement.

5.3. American Options Valuation

One unique benefit for the lattice approach is its ability to value American options. We use the proposed lattice model (Best-Fit) to value American put option under the Heston SV model with parameters: strike $K = \$100$, $\lambda = 3.0$, $m = 0.04$, $r = 0.05$, $q = 0$, and $\zeta = 0.1$. The results are summarized in Table 6. We compare the result with that reported in [Beliaeva and Nawalkha \(2010\)](#), denoted by ‘B&N Tree’ in the same table. We also apply the control variate (CV) technique to value the American put as follows:

$$\text{American Put (CV)} = \text{American Put (Best-Fit)} + \left(\text{European Put (Closed-Form)} - \text{European Put (Best-Fit)} \right)$$

Each of the last two columns of Table 6 indicates the difference, labeled as ‘error’, between an obtained American put option prices (by either the proposed lattice model or the B&N Tree) and that by the CV technique. It can be seen that most errors are smaller than one cent, and, in most cases, the differences obtained by the Best-Fit model are smaller than the corresponding ones reported in [Beliaeva and Nawalkha \(2010\)](#). Furthermore, the approach that achieves a smaller error is highlighted in bold in each comparison.

A more detailed comparison is summarized in Table 7. There are 36 cases tested in Table 6 (corresponding to 36 rows), which can be divided into several categories given in Table 7. The number of wins (in terms of a lower error) and the percentage of winning are recorded, along with the corresponding RMSE of each category. Overall, the Best-Fit model achieves smaller errors in 61% of all 36 test cases; our RMSE (0.0081) is only 44% of that (0.0181) by the B&N Tree (see Table 6). In all categories summarized in Table 7, the Best-Fit model outperforms the B&N tree approach in all categories either by the number of case wins or RMSE. In terms of the correlation ρ , Table 7 shows that the Best-Fit model performs especially well when the correlation is high ($\rho = -0.7$) such that the Best-Fit model does better in 67% of cases, and the RMSE is only 1/3 of the B&N Tree approach. When $|\rho|$ is high, lattice feasibility becomes harder to meet. Since the proposed lattice model maintains lattice feasibility and can best fit the underlying probability distribution for all ρ , the results in Table 7 show that it indeed performs relatively well when $|\rho|$ is high.

To summarize Table 7, it is fair to say that the Best-Fit model performs better especially when $|\rho|$ is high, T is greater than one month, and/or for the options that are ITM or OTM.

Table 6. American put price computed using Best-Fit model and control variate (CV) technique, compared with the result reported in [Beliaeva and Nawalkha \(2010\)](#).

S_0	ρ	V_0	T	Best-Fit ($N = 50$)	Best-Fit CV ($N = 50$)	B&N Tree ($N = 50$)	B&N Tree CV ($N = 50$)	Euro Put True Value	Error (Best-Fit)	Error (B&N Tree)	
90	-0.1	0.04	0.0833	10.0000	9.9996	10.0000	10.0001	9.6699	0.0004	0.0001	
90	-0.7	0.04	0.0833	10.0000	10.0010	10.0000	9.9996	9.6533	0.0010	0.0004	
100	-0.1	0.04	0.0833	2.1253	2.1262	2.1257	2.1249	2.0950	0.0009	0.0008	
100	-0.7	0.04	0.0833	2.1279	2.1274	2.1250	2.1263	2.0971	0.0006	0.0013	
110	-0.1	0.04	0.0833	0.1092	0.1090	0.1087	0.1090	0.1083	0.0002	0.0003	
110	-0.7	0.04	0.0833	0.1259	0.1275	0.1282	0.1273	0.1265	0.0016	0.0009	
90	-0.1	0.16	0.0833	10.7103	10.7109	10.7195	10.7099	10.5957	0.0006	0.0096	
90	-0.7	0.16	0.0833	10.6858	10.6829	10.6709	10.6800	10.5668	0.0030	0.0091	
100	-0.1	0.16	0.0833	4.2187	4.2166	4.2275	4.2163	4.1859	0.0021	0.0112	
100	-0.7	0.16	0.0833	4.2196	4.2156	4.2184	4.2147	4.1852	0.0040	0.0037	
110	-0.1	0.16	0.0833	1.1721	1.1667	1.1667	1.1669	1.1608	0.0053	0.0002	
110	-0.7	0.16	0.0833	1.1962	1.1944	1.2176	1.1942	1.1882	0.0018	0.0234	
90	-0.1	0.04	0.2500	10.1653	10.1654	10.1678	10.1663	9.6430	0.0002	0.0015	
90	-0.7	0.04	0.2500	10.1197	10.1208	10.1162	10.1207	9.5698	0.0010	0.0045	
100	-0.1	0.04	0.2500	3.4750	4.4765	3.4736	3.4727	3.3684	0.0014	0.0009	
100	-0.7	0.04	0.2500	3.4848	3.4816	3.4803	3.4801	3.3770	0.0032	0.0002	
110	-0.1	0.04	0.2500	0.7725	0.7729	0.7722	0.7730	0.7584	0.0003	0.0008	
110	-0.7	0.04	0.2500	0.8429	0.8417	0.8462	0.8412	0.8259	0.0012	0.0050	
90	-0.1	0.16	0.2500	12.1895	12.1773	12.1898	12.1752	11.8933	0.0122	0.0146	
90	-0.7	0.16	0.2500	12.1256	12.1135	12.0626	12.1061	11.8287	0.0121	0.0435	
100	-0.1	0.16	0.2500	6.5014	6.4937	6.5027	6.4936	6.3755	0.0077	0.0091	
100	-0.7	0.16	0.2500	6.4986	6.4898	6.4855	6.4908	6.3735	0.0089	0.0053	
110	-0.1	0.16	0.2500	3.0985	3.0900	3.0874	3.0900	3.0451	0.0085	0.0026	
110	-0.7	0.16	0.2500	3.1550	3.1453	3.1731	3.1479	3.1011	0.0097	0.0252	
90	-0.1	0.04	0.5000	10.6472	10.6466	10.6431	10.6422	9.8582	0.0006	0.0009	
90	-0.7	0.04	0.5000	10.5627	10.5636	10.5575	10.5609	9.7572	0.0009	0.0034	
100	-0.1	0.04	0.5000	4.6481	4.6492	4.6445	4.6439	4.4126	0.0011	0.0006	
100	-0.7	0.04	0.5000	4.6685	4.6636	4.6672	4.6610	4.4312	0.0050	0.0062	
110	-0.1	0.04	0.5000	1.6812	1.6812	1.6804	1.6813	1.6220	0.0001	0.0009	
110	-0.7	0.04	0.5000	1.7906	1.7872	1.7950	1.7856	1.7240	0.0034	0.0094	
90	-0.1	0.16	0.5000	13.3245	13.3089	13.2938	13.2944	12.7057	0.0155	0.0006	
90	-0.7	0.16	0.5000	13.2393	13.2230	13.0958	13.1731	12.6171	0.0162	0.0773	
100	-0.1	0.16	0.5000	8.0210	8.0041	7.9928	7.9943	7.6974	0.0168	0.0015	
100	-0.7	0.16	0.5000	8.0174	8.0001	7.9432	7.9810	7.6965	0.0174	0.0378	
110	-0.1	0.16	0.5000	4.5590	4.5430	4.5221	4.5383	4.3942	0.0160	0.0162	
110	-0.7	0.16	0.5000	4.6343	4.6188	4.5972	4.6118	4.4716	0.0155	0.0146	
									RMSE	0.0080	0.0181

Table 7. Comparison summary of the results in Table 6.

Category	No. of Cases	No. of Wins (%)		RMSE	
		Best-Fit	B&N Tree	Best-Fit	B&N Tree
Overall	36	22 (61%)	14 (39%)	0.0080	0.0181
$\rho = -0.1$	18	10 (56%)	8 (44%)	0.0078	0.0066
$\rho = -0.7$	18	12 (67%)	6 (33%)	0.0082	0.0247
$V_0 > m$	18	11 (61%)	7 (39%)	0.0112	0.0254
$V_0 \leq m$	18	11 (61%)	7 (39%)	0.0018	0.0033
In-the-money (ITM)	12	9 (75%)	3 (25%)	0.0082	0.0263
At-the-money (ATM)	12	5 (42%)	7 (58%)	0.0081	0.0120
Out-of-the-money (OTM)	12	8 (67%)	4 (33%)	0.0077	0.0122
$T = 1/12$ yr	12	6 (50%)	6 (50%)	0.0023	0.0085
$T = 1/4$ yr	12	8 (67%)	4 (33%)	0.0072	0.0156
$T = 1/2$ yr	12	8 (67%)	4 (33%)	0.0116	0.0259

6. Conclusions

In this paper, we focus on two-factor lattices for general diffusion processes where volatilities can be state-dependent, including stochastic volatility models. For a lattice approach, although it is common knowledge that branching probabilities must be between zero and one, few methods can guarantee all branching probabilities of all nodes in all stages are always legitimate. We refer to this property as lattice feasibility. Since it is not unusual to encounter negative probabilities, some practitioners have argued that negative

probabilities are not necessarily ‘bad’ and may be further exploited. We have developed a theoretical framework of lattice feasibility to investigate how negative probabilities may impact option pricing in a lattice approach. We have shown that lattice feasibility can be achieved by adjusting a lattice’s configuration (e.g., grid sizes).

Failing to meet lattice feasibility means that some branching probabilities in a lattice are negative or exceed unity, which implies distortions on the probability distribution of the underlying variables. Depending on the distortion, the accuracy of options pricing may be affected. We have found that out-of-the-money options are most affected, followed by in-the-money options and at-the-money options. It has also been observed that negative probabilities indeed matter less in some situations. Since, in reality, how an underlying probability distribution is distorted by lattice infeasibility is unknown a priori, it seems unlikely that one could exploit negative probabilities consistently as some practitioners may hope.

Although lattice feasibility is a necessary condition for weak convergence of approximating the underlying diffusion processes, our numerical tests also show that lattice feasibility alone may not be sufficient to guarantee accurate valuation, especially when the time step of the lattice is not especially small. Since legitimate branching probabilities may not be unique, we use an optimization approach to find branching probabilities that are not only legitimate but also can best fit the probability distribution of the underlying variables. Extensive numerical tests show that this optimized lattice model is a reliable and robust approach for financial option valuations.

Author Contributions: Conceptualization, C.-L.T.; methodology, C.-L.T. and D.W.-C.M.; software, C.-L.T.; analysis, C.-L.T., D.W.-C.M., S.-L.C. and P.-T.S.; writing—original draft preparation, C.-L.T.; writing—review and editing, C.-L.T., D.W.-C.M., S.-L.C. and P.-T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Proof of Proposition 1

From (15a)–(15c), apparently

$$p_u \geq 0, p_d \geq 0, p_m \geq 0 \Leftrightarrow -\left(\frac{\epsilon}{h}\right)^2 + \frac{|\epsilon|}{h} \leq \frac{x^2}{c^2h^2} \leq 1 - \left(\frac{\epsilon}{h}\right)^2 \tag{A1}$$

For p_u , p_m , and p_d to be legitimate for all $\epsilon \in [-0.5, 0.5]$, we consider

$$\max_{|\epsilon| \leq 0.5} \left(-\left(\frac{\epsilon}{h}\right)^2 + \frac{|\epsilon|}{h} \right) \leq \frac{x^2}{c^2h^2} \leq \min_{|\epsilon| \leq 0.5} \left(1 - \left(\frac{\epsilon}{h}\right)^2 \right). \tag{A2}$$

The upper bound of (A2) is $1 - 1/(2h^2)$ achieved at $\epsilon = 0.5$. The lower bound is achieved at $\epsilon = \pm 0.5$. Thus,

$$\frac{h}{2} - \frac{1}{4} \leq \left(\frac{x}{c}\right)^2 \leq h^2 - \frac{1}{4}, \tag{A3}$$

From (9), it is clear that the middle term of (A3) is within the following interval:

$$\left(\frac{x}{c}\right)^2 \in \begin{cases} [1/c^2, 1.5^2/c^2], & \text{if } h = 1, \\ [(h - 0.5)^2/c^2, (h + 0.5)^2/c^2], & \text{if } h \geq 2. \end{cases} \tag{A4}$$

For both (A3) and (A4) to hold, we must have for $h \in \mathbb{N}$ and

$$\frac{(h + 0.5)^2}{c^2} \leq h^2 - \frac{1}{4} \tag{A5a}$$

$$\max\left(\frac{1}{c^2}, \frac{(h - 0.5)^2}{c^2}\right) \geq \frac{h}{2} - \frac{1}{4}. \tag{A5b}$$

Thus,

$$\frac{h + 0.5}{h - 0.5} \leq c^2 \leq \max(1, (h - 0.5)^2) \left(\frac{h}{2} - \frac{1}{4}\right)^{-1} \tag{A6}$$

Plugging $h = 1, 2, 3, \dots$ to (A6) gives the following ranges of c^2 :

$$\begin{aligned} 3 \leq c^2 \leq 4, & \quad \text{if } h = 1 \\ 5/3 \leq c^2 \leq 3, & \quad \text{if } h = 2 \\ 7/5 \leq c^2 \leq 5, & \quad \text{if } h = 3 \\ 9/7 \leq c^2 \leq 7, & \quad \text{if } h = 4 \\ & \dots \end{aligned}$$

When $\underline{h} = 1$, apparently c^2 must be 3. When $h \geq 2$, the upper bound of c^2 increases with h , while the lower bound decreases. Therefore, for $h \geq \underline{h}$, \underline{h} sets the bounds. That is, $\sqrt{(\underline{h} + 0.5)/(\underline{h} - 0.5)} \leq c \leq \sqrt{2\underline{h} - 1}$ for $\underline{h} \geq 2$.

Appendix A.2. Proof of Proposition 2

The proof is based on Durrett (1996) where the conditions for weak convergence from Markov chains to diffusion processes are given. To present our proof, we first give the following lemma which is taken from Lemma 8.2 of (Durrett 1996, p. 306) and adapted to our one-factor case. For convenience, we introduce the following definitions for the one-factor lattice $Y_t^{\Delta t}$.

$$\begin{aligned} \alpha^{\Delta t}(y) &= (\Delta t)^{-1} \mathbb{E}[(Y_{t+\Delta t}^{\Delta t} - y) | Y_t^{\Delta t} = y], \\ \beta^{\Delta t}(y) &= (\Delta t)^{-1} \mathbb{E}[(Y_{t+\Delta t}^{\Delta t} - y)^2 | Y_t^{\Delta t} = y], \\ \gamma_p^{\Delta t}(y) &= (\Delta t)^{-1} \mathbb{E}[|Y_{t+\Delta t}^{\Delta t} - y|^p | Y_t^{\Delta t} = y]. \end{aligned}$$

They are concerned with the first, second, and higher (absolute) moments of the change in lattice across Δt .

Lemma A1. Suppose that $Y_0^{\Delta t} = Y_0$. If there exists a $p \geq 2$ and for all $R < \infty$, we have

$$\begin{aligned} (i) \quad & \limsup_{\Delta t \downarrow 0} \sup_{|y| \leq R} |\alpha^{\Delta t}(y) - \mu(y)| = 0, \\ (ii) \quad & \limsup_{\Delta t \downarrow 0} \sup_{|y| \leq R} |\beta^{\Delta t}(y) - \sigma^2(y)| = 0, \\ (iii) \quad & \limsup_{\Delta t \downarrow 0} \sup_{|y| \leq R} \gamma_p^{\Delta t}(y) = 0, \end{aligned}$$

then the one-factor lattice $Y_t^{\Delta t}$ converges weakly to Y_t as $\Delta t \rightarrow 0$.

To prove Proposition 2, we need to check the above three conditions. Note that conditions (i) and (ii) are true because (17a)–(17c) hold for all y . Then, we check condition (iii) with $p = 4$ and have

$$\begin{aligned} \gamma_4^{\Delta t}(y) &= (\Delta t)^{-1} E[|Y_{t+\Delta t}^{\Delta t} - y|^4 | Y_t^{\Delta t} = y] \\ &= (\Delta t)^{-1} (p_u |(k + h)\Delta y|^4 + p_m |k\Delta y|^4 + p_d |(k - h)\Delta y|^4) \\ &\leq (\Delta t)^{-1} (|(k + h)\Delta y|^4 + |k\Delta y|^4 + |(k - h)\Delta y|^4) \\ &\leq (\Delta t)^{-1} 3(k + h)^4 (\Delta y)^4. \end{aligned}$$

Note that

$$\Delta y = c\sigma^s \sqrt{\Delta t}, \quad k(y) = \left\lfloor \frac{\mu(y)\Delta t}{\Delta y} + 0.5 \right\rfloor, \quad h(y) = \left\lfloor \frac{\sigma(y)}{\sigma^s} + 0.5 \right\rfloor.$$

Fix an arbitrarily large $R < \infty$. For any y such that $|y| \leq R$ (thus $\mu(y)$ and $\sigma(y)$ are finite) and for a sufficiently small Δt , we have

$$|k(y)| \leq \sup_{|y| \leq R} \left| \frac{\mu(y)}{c\sigma^s} \sqrt{\Delta t} + 0.5 \right| \leq 1, \quad |h(y)| \leq \sup_{|y| \leq R} \left| \frac{\sigma(y)}{\sigma^s} \right| + 1 = M < \infty.$$

Putting these together, we have

$$\sup_{|y| \leq R} \gamma_4^{\Delta t}(y) \leq 3c^4 (\sigma^s)^4 (M + 1)^4 \Delta t \rightarrow 0 \quad \text{as } \Delta t \rightarrow 0.$$

Thus, condition (iii) is satisfied.

Appendix A.3. Proof of Proposition 3

Before we prove Proposition 3, some preparation needs to be done. Equation (20) is equivalent to:

$$\begin{aligned} &\left\lfloor \frac{y}{\Delta y} \right\rfloor + \left\lfloor \frac{-\lambda(y-m)\Delta t}{\Delta y} + 0.5 \right\rfloor - \left\lfloor \frac{\xi\sqrt{y}}{\sigma^s} + 0.5 \right\rfloor - \left\lfloor \frac{y^{\min}}{\Delta y} \right\rfloor \\ &= \left\lfloor \frac{y}{\Delta y} \right\rfloor + \left\lfloor \frac{-\lambda(y-m)\Delta t}{\Delta y} + 0.5 \right\rfloor - \left\lfloor \frac{\xi c \sqrt{\Delta t} \sqrt{y}}{\Delta y} + 0.5 \right\rfloor - \left\lfloor \frac{y^{\min}}{\Delta y} \right\rfloor \geq 0 \end{aligned} \tag{A7a}$$

Likewise, (21) is equivalent to:

$$\left\lfloor \frac{-\lambda(y-m)\Delta t}{\Delta y} + 0.5 \right\rfloor - \left\lfloor \frac{\xi c \sqrt{\Delta t} \sqrt{y}}{\Delta y} + 0.5 \right\rfloor \geq 0 \tag{A7b}$$

Since dealing with the floor functions is cumbersome, the following lemma enables us to consider the sufficient conditions without the floor functions that imply (A7a) and (A7b).

Lemma A2. Suppose $a_1, \dots, a_4 \in \mathbb{R}$. The following two statements are true:

- (a) If $a_1 + a_2 - a_3 - a_4 \geq 1$, then $\lfloor a_1 \rfloor + \lfloor a_2 + 0.5 \rfloor - \lfloor a_3 + 0.5 \rfloor - \lfloor a_4 \rfloor \geq 0$.
- (b) If $a_2 - a_3 \geq 0$, then $\lfloor a_2 + 0.5 \rfloor - \lfloor a_3 + 0.5 \rfloor \geq 0$.

Proof. Let $n_1 = \lfloor a_1 \rfloor$, $n_2 = \lfloor a_2 + 0.5 \rfloor$, $n_3 = \lfloor a_3 + 0.5 \rfloor$, and $n_4 = \lfloor a_4 \rfloor$. This implies that $a_1 \in [n_1, n_1 + 1)$, $a_2 \in [n_2 - 0.5, n_2 + 0.5)$, $a_3 \in [n_3 - 0.5, n_3 + 0.5)$, and $a_4 \in [n_4, n_4 + 1)$.

- (a) We have $a_1 + a_2 - a_3 - a_4 \in (n_1 + n_2 - n_3 - n_4 - 2, n_1 + n_2 - n_3 - n_4 + 2)$. If $n_1 + n_2 - n_3 - n_4 < 0$, i.e., $n_1 + n_2 - n_3 - n_4 \leq -1$, then $a_1 + a_2 - a_3 - a_4 < 1$. This means that $a_1 + a_2 - a_3 - a_4 \geq 1$ implies $n_1 + n_2 - n_3 - n_4 \geq 0$.

- (b) Similarly, $a_2 - a_3 \in (n_2 - n_3 - 1, n_2 - n_3 + 1)$. If $n_2 - n_3 < 0$, i.e., $n_2 - n_3 \leq -1$, then $a_2 - a_3 < 0$. This means that, if $a_2 - a_3 \geq 0$, then $n_2 - n_3 \geq 0$.
□

Using Lemma A2, we consider the sufficient condition of (A7a) as follows:

$$\frac{y}{\Delta y} + \frac{-\lambda(y - m)\Delta t}{\Delta y} - \frac{\xi c \sqrt{\Delta t} \sqrt{y}}{\Delta y} - \frac{y^{\min}}{\Delta y} \geq 1, \tag{A8}$$

which is equivalent to

$$(1 - \lambda\Delta t)y - \xi c \sqrt{\Delta t} \sqrt{y} + (\lambda m \Delta t - y^{\min} - \Delta y) \geq 0. \tag{A9}$$

We want to show that $y^{\min} > 0$ exists such that (A9) holds for all $y \geq 0$. Then, using Lemma A2(a), this implies that (20) holds for all $y \geq 0$.

For (A7b), we consider the following sufficient condition using Lemma A2(b):

$$\frac{\lambda(m - \Delta y)\Delta t}{\Delta y} - \frac{\xi c \sqrt{\Delta y} \sqrt{\Delta t}}{\Delta y} = \frac{-\Delta t}{\Delta y} (\Delta y + \frac{\xi c}{\sqrt{\Delta t}} \sqrt{\Delta y} - \lambda m) \geq 0, \tag{A10}$$

which is equivalent to

$$\Delta y + \frac{\xi c}{\sqrt{\Delta t}} \sqrt{\Delta y} \leq \lambda m. \tag{A11}$$

Lemma A3. *The following two statements are true:*

- (a) Given $a, b, d \in \mathbb{R}$, if $a > 0, b > 0, d > 0$, and $b^2 \leq 4ad$, then $ay - b\sqrt{y} + d \geq 0$ for $\forall y \geq 0$.
 (b) If $\alpha_1 \geq 0$ and $\alpha_2 > 0$, then $y + \alpha_1 \sqrt{y} \leq \alpha_2$ has a solution $y > 0$.

Proof.

- (a) Given $a, b, d \in \mathbb{R}, a > 0, b > 0, d > 0$, and $b^2 \leq 4ad$, consider $f(y) = a^2y^2 + (2ad - b^2)y + d^2$. Note that $f(y)$ is a quadratic convex function. Consider two cases: (I) If $b^2 \leq 2ad$, then $f'(0) = 2ad - b^2 \geq 0$; $f(y)$ is increasing and is positive for $\forall y \geq 0$. (II) If $2ad < b^2 \leq 4ad$, f has a local minimum at $y^* = (b^2 - 2ad)/(2a^2) > 0$ with objective value $f(y^*) = b^2(4ad - b^2)/(4a^2) \geq 0$. Both cases imply that $f(y) \geq 0, \forall y \geq 0$. Since $a > 0$ and $d > 0$, $f(y) = a^2y^2 + (2ad - b^2)y + d^2 \geq 0, \forall y \geq 0$ implies that $ay + d \geq b\sqrt{y}, \forall y \geq 0$.

- (b) When $y = 0$, the left-hand-side of $y + \alpha_1 \sqrt{y} = 0 < \alpha_2$. Therefore, by continuity, there exists some $y > 0$ to satisfy the inequality $y + \alpha_1 \sqrt{y} \leq \alpha_2$.
□

Now, we are ready to prove Proposition 3 using Lemmas A2 and A3. Let $a = 1 - \lambda\Delta t, b = \xi c \sqrt{\Delta t}, d = \lambda m \Delta t - y_{\min} - \Delta y$, where Δy is from (19). From Lemma A3(a), if $a = 1 - \lambda\Delta t > 0$, which is condition (i) of Proposition 3, and $b^2 \leq 4ad$, then (A9) holds for $\forall y \geq 0$, which further implies that (20) holds for $\forall y \geq 0$. Now, consider the following equivalent statements:

$$b^2 \leq 4ad \tag{A12}$$

$$\Leftrightarrow \xi^2 c^2 \Delta t \leq 4(1 - \lambda\Delta t)(\lambda m \Delta t - y_{\min} - \frac{\xi c}{\max(\underline{h} - 0.5, 1)} \sqrt{y_{\min}} \sqrt{\Delta t}) \tag{A13}$$

$$\Leftrightarrow y^{\min} + \frac{\xi c \sqrt{\Delta t}}{\max(\underline{h} - 0.5, 1)} \sqrt{y^{\min}} \leq \Delta t \left(\lambda m - \frac{\xi^2 c^2}{4(1 - \lambda\Delta t)} \right). \tag{A14}$$

Using Lemma A3(b), $y^{\min} > 0$ exists if the right-hand side of (A14) is strictly positive. That is,

$$\lambda m > \frac{\xi^2 c^2}{4(1 - \lambda \Delta t)}, \tag{A15}$$

which is given by condition (ii) of Proposition 3. Suppose a feasible $y^{\min} = \tilde{y}_a > 0$ is found that satisfies (A14). Using Lemma A2(a) and Lemma A3(a), we conclude $y^{\min} > 0$ exists such that (20) holds for $\forall y \geq 0$.

Using $y^{\min} = \tilde{y}_a$ to evaluate Δy from (19), if $\tilde{y}_a \geq \Delta y$, we are done. Otherwise, $k(\Delta y) - h(\Delta y) \geq 0$ is imposed as given in (21), which is reduced to (A11). Using Lemma A3(b), it is clear that $\Delta y > 0$ exists to ensure (A11). However, to ensure that $y^{\min} < \Delta y$, from (19), the following upper bound of Δy must also be imposed:

$$\Delta y < \xi^2 c^2 \Delta t / \max(\underline{h} - 0.5, 1)^2 \tag{A16}$$

Using Lemma A3(b) again, it is clear that a $\Delta y > 0$ exists that also meets (A16). Since Δy is a function of y^{\min} from (19), we conclude that a $y^{\min} > 0$ exists, say $y^{\min} = \tilde{y}_b$, such that (21) holds. Since $\tilde{y}_b < \tilde{y}_a$, it is clear that \tilde{y}_a also meets (A14). Therefore, both (A7a) and (A7b) hold with $y^{\min} = \tilde{y}_b$.

To sum up, given conditions (i) and (ii) of Proposition 3, $y^{\min} > 0$ exists (e.g., equal to $\min(\tilde{y}_a, \tilde{y}_b) > 0$). This means that the entire lattice can be well contained in \mathbb{R}^+ .

Appendix A.4. Proof of Proposition 4

Given Feller’s condition $\lambda m \geq 0.5\xi^2$, we want to show there exists a lattice configuration $(\Delta t, \underline{h}, c)$ such that $c^2 / (1 - \lambda \Delta t) \leq 2$. To do so, we choose a small $\Delta t > 0$ such that $\lambda \Delta t < 0.5$, i.e., $2(1 - \lambda \Delta t) > 1$. We also choose a big \underline{h} such that c^2 is very close to 1 while still being greater than 1 based on (16), and furthermore $2(1 - \lambda \Delta t) \geq c^2 > 1$. This would imply $1 - \lambda \Delta t > 0$, which is condition (i) of Proposition 3; and

$$\lambda m \geq 0.5\xi^2 \geq \frac{c^2}{4(1 - \lambda \Delta t)} \xi^2, \tag{A17}$$

which is condition (ii) of Proposition 3. The result follows from Proposition 3.

Appendix A.5. Proof of Proposition 5

The proof for the two-factor case is in the same spirit as the one-factor case treated in Section A.2. For the two-factor lattice $(Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t})$, define

$$\begin{aligned} \alpha_l^{\Delta t}(y_1, y_2) &= (\Delta t)^{-1} \mathbb{E}[(Y_{l,t+\Delta t}^{\Delta t} - y_l) | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)], \quad l = 1, 2, \\ \beta_l^{\Delta t}(y_1, y_2) &= (\Delta t)^{-1} \mathbb{E}[(Y_{l,t+\Delta t}^{\Delta t} - y_l)^2 | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)], \quad l = 1, 2, \\ \delta^{\Delta t}(y_1, y_2) &= (\Delta t)^{-1} \mathbb{E}[(Y_{1,t+\Delta t}^{\Delta t} - y_1)(Y_{2,t+\Delta t}^{\Delta t} - y_2) | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)], \\ \gamma_p^{\Delta t}(y_1, y_2) &= (\Delta t)^{-1} \mathbb{E}[(Y_{1,t+\Delta t}^{\Delta t} - y_1)^2 + (Y_{2,t+\Delta t}^{\Delta t} - y_2)^2 | (Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t}) = (y_1, y_2)]. \end{aligned}$$

With the above definitions, we may restate Lemma 8.2 of Durrett (1996) for the two-factor case as follows.

Lemma A4. Suppose that $(Y_{1,0}^{\Delta t}, Y_{2,0}^{\Delta t}) = (Y_{1,0}, Y_{2,0})$. If there exists a $p \geq 2$ and for all $R < \infty$, we have

$$\begin{aligned}
 (i) \quad & \lim_{\Delta t \downarrow 0} \sup_{|(y_1, y_2)| \leq R} |\alpha_l^{\Delta t}(y_1, y_2) - \mu_l(y_1, y_2)| = 0, \quad l = 1, 2, \\
 (ii) \quad & \lim_{\Delta t \downarrow 0} \sup_{|(y_1, y_2)| \leq R} |\beta_l^{\Delta t}(y_1, y_2) - \sigma_l^2(y_1, y_2)| = 0, \quad l = 1, 2, \\
 & \lim_{\Delta t \downarrow 0} \sup_{|(y_1, y_2)| \leq R} |\delta^{\Delta t}(y_1, y_2) - \rho(y_1, y_2)\sigma_1(y_1, y_2)\sigma_2(y_1, y_2)| = 0, \\
 (iii) \quad & \lim_{\Delta t \downarrow 0} \sup_{|(y_1, y_2)| \leq R} \gamma_p^{\Delta t}(y_1, y_2) = 0,
 \end{aligned}$$

then $(Y_{1,t}^{\Delta t}, Y_{2,t}^{\Delta t})$ converges weakly to $(Y_{1,t}, Y_{2,t})$ as $\Delta t \rightarrow 0$.

The proof of Proposition 5 requires the above three conditions. Conditions (i) and (ii) are true because (25a)–(25d) hold. To check condition (iii) with $p = 4$, we observe that

$$\begin{aligned}
 \gamma_4^{\Delta t}(y_1, y_2) &= (\Delta t)^{-1} \sum_{i,j \in \Omega} p_{ij} [((k_1 + n_1 h_1)\Delta y_1)^2 + ((k_2 + n_2 h_2)\Delta y_2)^2]^2 \\
 &\leq (\Delta t)^{-1} 9 [((k_1 + h_1)\Delta y_1)^2 + ((k_2 + h_2)\Delta y_2)^2]^2,
 \end{aligned}$$

where $n_1, n_2 \in \{-1, 0, 1\}$. Using an argument similar to Section A.2, we conclude that, for a given $R < \infty$, there exists a constant $M < \infty$ such that $\gamma_4^{\Delta t}(y_1, y_2) \leq M\Delta t$ for all lattice nodes with $|(y_1, y_2)| < R$. This shows that condition (iii) is satisfied.

Appendix A.6. Proof of Proposition 6

(i) First, observe from (15a)–(15c) that replacing ϵ by $-\epsilon$ is equivalent to switching p_u and p_d . Consider the optimization problem of R^{\max} with a feasible solution of $\{p_{ij}\}, i, j \in \Omega$, as shown in Figure A1. The objective function is $p_{uu} + p_{dd} - p_{ud} - p_{du}$, which is the sum of the northwest and southeast corner elements minus the sum of the other two corner elements. It can be seen that switching the first and the third columns (i.e., replacing \tilde{p}_{2u} and \tilde{p}_{2d}), and the first and the third rows (i.e., replacing \tilde{p}_{1u} and \tilde{p}_{1d}), yields the same objective value. This implies that $R^{\max}(\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2) = R^{\max}(-\epsilon_1, x_1, \gamma_1, h_1, -\epsilon_2, x_2, \gamma_2, h_2)$.

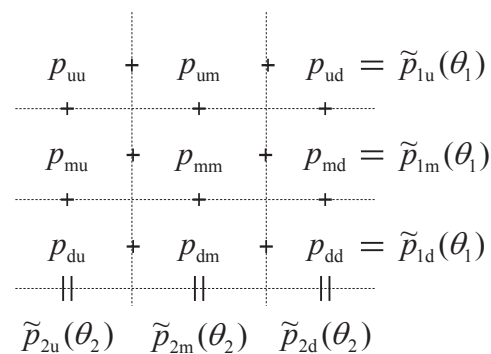


Figure A1. Interpretation of (Q).

(ii) Continuing the argument in (i), if one only switches the first and the third columns or the first and the third rows, the objective value will still be the same but with a different sign. Therefore, $R^{\max}(\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2) = -R^{\min}(\epsilon_1, x_1, \gamma_1, h_1, -\epsilon_2, x_2, \gamma_2, h_2) = -R^{\min}(-\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2)$.

(iii) From (ii), it is clear that

$$\min_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(R^{\max}(\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \tag{A18a}$$

$$= \min_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(-R^{\min}(-\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \tag{A18b}$$

$$= - \max_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(R^{\min}(-\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2) - \frac{(-\epsilon_1) \epsilon_2}{h_1 h_2} \right) \tag{A18c}$$

$$= - \max_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(R^{\min}(\epsilon_1, x_1, \gamma_1, h_1, \epsilon_2, x_2, \gamma_2, h_2) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \tag{A18d}$$

In (A18d), we use the fact that $\epsilon_1 \in [-0.5, 0.5]$; thus, replacing $-\epsilon_1$ by ϵ_1 will not affect the result of the optimization.

Appendix A.7. Proof of Proposition 7

See the proof of Lemma 2 in Tseng and Lin (2007).

Appendix A.8. Proof of Proposition 8

Let $f_0(z) \equiv \min(f_1(z), \dots, f_n(z))$, and $z^i \equiv \arg \min_{z \in D} f_i(z)$, $i = 0, \dots, n$. For simplicity, we assume $f_1(z^1) \leq f_i(z^i), \forall i = 1, \dots, n$. We want to show $f_0(z^0) = f_1(z^1)$.

By the definition of $z^i, i = 1, \dots, n$, we have $f_1(z^1) = \min\{f_1(z^1), \dots, f_n(z^1)\} = f_0(z^1)$. Thus, $f_1(z^1) \geq f_0(z^0)$. Furthermore, we have $f_i(z^i) \leq f_i(z^0), i = 1, \dots, n$. Thus, $\min\{f_1(z^1), \dots, f_n(z^1)\} \leq \min\{f_1(z^0), \dots, f_n(z^0)\}$. This means that $f_1(z^1) \leq f_0(z^0)$. Thus, we conclude that $f_0(z^0) = f_1(z^1)$.

Appendix A.9. Proof of Theorem 1

We consider the six cases of $R_i^{\max}, i = 1, \dots, 6$, given in Proposition 7.

(i) $R_1^{\max} = \tilde{p}_{1u} + \tilde{p}_{1d}$. From (31), we consider the following optimization problem:

$$\min_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(R_1^{\max} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \tag{A19a}$$

$$= \min_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(\frac{\gamma_1^2}{c_1^2} + \frac{\epsilon_1^2}{h_1^2} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \tag{A19b}$$

$$= \min_{\theta_l \in \Theta_l(h_l)} \frac{1}{\gamma_1 \gamma_2} \left(\frac{\gamma_1^2}{c_1^2} + \frac{\phi_1^2}{h_1^2} - \frac{\phi_1}{2h_1 h_2} \right) \tag{A19c}$$

$$= w_1, \tag{A19d}$$

where, in (A19c), we optimize the objective function over ϵ_1 and ϵ_2 first, with $\gamma_l > 0$ and $h_l > 0$ being fixed. The optimal solution is achieved at $(\epsilon_1, \epsilon_2) = (\phi_1, 1/2)$, with $\phi_1 = \min(h_1 / (4h_2), 1/2)$.

(ii) This is a symmetric case for case (i) by exchanging the factor indices, 1 and 2.

(iii) $R_3^{\max} = \tilde{p}_{1u} + \tilde{p}_{2d}$, we have

$$\min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} (R_3^{\max} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2}) \tag{A20a}$$

$$= \min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} \left[\frac{1}{2} \left(\frac{\gamma_1^2}{c_1^2} + \frac{\epsilon_1^2}{h_1^2} + \frac{\epsilon_1}{h_1} \right) + \frac{1}{2} \left(\frac{\gamma_2^2}{c_2^2} + \frac{\epsilon_2^2}{h_2^2} - \frac{\epsilon_2}{h_2} \right) - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right] \tag{A20b}$$

$$= \min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} \left[\frac{1}{2} \left(\left(\frac{\epsilon_1}{h_1} - \frac{\epsilon_2}{h_2} + \frac{1}{2} \right)^2 - \frac{1}{4} \right) + \frac{1}{2} \left(\frac{\gamma_1^2}{c_1^2} + \frac{\gamma_2^2}{c_2^2} \right) \right] \tag{A20c}$$

$$= \min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} \left[\frac{1}{8} (\phi_3^2 - 1) + \frac{1}{2} \left(\frac{\gamma_1^2}{c_1^2} + \frac{\gamma_2^2}{c_2^2} \right) \right] \tag{A20d}$$

$$= w_3. \tag{A20e}$$

Again, in (A20d), we optimize over (ϵ_1, ϵ_2) first, and the optimal solution is achieved at $(\epsilon_1, \epsilon_2) = (-0.5, 0.5)$ with $\phi_3 = \max\{1 - 1/h_1 - 1/h_2, 0\}$.

(iv) $R_4^{\max} = 1 - (\tilde{p}_{1d} - \tilde{p}_{2d}) - (\tilde{p}_{2u} - \tilde{p}_{1u})$. Repeat the same process and we have

$$\min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} (R_4^{\max} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2}) \tag{A21a}$$

$$= \min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} \left(1 + \frac{\epsilon_1}{h_1} - \frac{\epsilon_2}{h_2} - \frac{\epsilon_1 \epsilon_2}{h_1 h_2} \right) \tag{A21b}$$

$$= \min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} \left(1 + \frac{\epsilon_1}{h_1} \right) \left(1 - \frac{\epsilon_2}{h_2} \right) \tag{A21c}$$

$$= \min_{\theta_l \in \Theta_l(\underline{h}_l)} \frac{1}{\gamma_1 \gamma_2} \left(1 - \frac{1}{2h_1} \right) \left(1 - \frac{1}{2h_2} \right) \tag{A21d}$$

$$= w_4, \tag{A21e}$$

where in (A21d) we used the fact that the optimal solution for (ϵ_1, ϵ_2) is $(-1/2, 1/2)$.

(v) For cases associated with R_5^{\max} and R_6^{\max} , they are symmetric counterparts of (iii) and (iv) by exchanging the factor indices 1 and 2. This obtains the same lower bounds as in (iii) and (iv).

Summarizing all four possible cases above, we conclude that

$$\bar{\rho} = c_1 c_2 \cdot \min\{w_1, w_2, w_3, w_4\}. \tag{A22}$$

The proof is completed.

Appendix A.10. Proof of Proposition 9

From (11) and (40), we have $\sigma_1^s = \sigma_2^s / \xi$. It follows that at any node

$$x_1 = \frac{\sqrt{y_2}}{\sigma_1^s} = \frac{\xi \sqrt{y_2}}{\sigma_2^s} = x_2, \tag{A23}$$

and

$$h_1 = \lfloor x_1 + 0.5 \rfloor = \lfloor x_2 + 0.5 \rfloor = h_2. \tag{A24}$$

Thus,

$$\gamma_1 = \frac{x_1}{h_1} = \frac{x_2}{h_2} = \gamma_2. \tag{A25}$$

Appendix A.11. Proof of Theorem 2

This proof is similar to that of Theorem 1. Six cases corresponding to R_1^{\max} to R_6^{\max} will be discussed.

(i) $R_1^{\max} = \tilde{p}_{1u} + \tilde{p}_{1d}$. Consider

$$\min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(R_1^{\max}(\theta_1, \theta_2) \right) = \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(\frac{x^2}{c_1^2 h^2} \right) = \frac{1}{c_1^2}$$

This corresponds to w_1^H .

(ii) $R_2^{\max} = \tilde{p}_{2u} + \tilde{p}_{2d}$. Consider

$$\min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(R_2^{\max}(\theta_1, \theta_2) \right) = \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(\frac{x^2}{c_2^2 h^2} + \frac{\epsilon_2^2}{h^2} \right) = \frac{1}{c_2^2},$$

where the minimum is achieved at $\epsilon_2 = 0$. This corresponds to w_2^H .

(iii) $R_3^{\max} = \tilde{p}_{1u} + \tilde{p}_{2d}$, we have

$$\begin{aligned} & \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(R_3^{\max}(\theta_1, \theta_2) \right) \\ &= \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(\frac{1}{2} \left(\frac{x^2}{c_1^2 h^2} + \frac{x^2}{c_2^2 h^2} \right) + \frac{1}{2} \left(-\frac{\epsilon_2}{h} + \frac{1}{2} \right)^2 - \frac{1}{8} \right) \end{aligned} \tag{A26}$$

$$= \frac{1}{2} \left(\frac{1}{c_1^2} + \frac{1}{c_2^2} \right) + \min_{x,h} \frac{h^2}{x^2} \left(\frac{1}{2} \left(-\frac{1}{2h} + \frac{1}{2} \right)^2 - \frac{1}{8} \right) \tag{A27}$$

$$= \frac{1}{2} \left(\frac{1}{c_1^2} + \frac{1}{c_2^2} \right) + \frac{\hat{h}^2}{(\hat{h} - 0.5)^2} \left(\frac{1}{2} \left(-\frac{1}{2\hat{h}} + \frac{1}{2} \right)^2 - \frac{1}{8} \right) \tag{A28}$$

where in the minimization of (A26) over ϵ_2 , the minimum is achieved at $\epsilon_2 = 0.5$. To solve the optimization in (A27), treat the objective function as a one-dimensional function of $x \geq 1$, which is a continuous variable, with h plugged in as $\lfloor x + 0.5 \rfloor$ from (9). It can be seen that the objective function is discontinuous. Using computing tools to display the one-dimensional function, subject to $x \geq \max(\underline{h} - 0.5, 1)$ from (14) and $h \geq \underline{h}$, the global minimum is achieved at $(h, x) = (\hat{h}, \hat{h} - 0.5)$, where $\hat{h} = \max(\underline{h}, 2)$.

(iv) $R_4^{\max} = 1 - (\tilde{p}_{1d} - \tilde{p}_{2d}) - (\tilde{p}_{2u} - \tilde{p}_{1u})$.

$$\min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(R_4^{\max}(\theta_1, \theta_2) \right) = \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(1 - \frac{\epsilon_2}{h} \right) \tag{A29}$$

$$= \min_{x,h} \frac{h^2}{x^2} \left(1 - \frac{1}{2h} \right) = \frac{\underline{h}^2}{(\underline{h} + 0.5)^2} \left(1 - \frac{1}{2\underline{h}} \right) \tag{A30}$$

where, in the minimization of (A29) over ϵ_2 , the minimum is achieved at $\epsilon_2 = 0.5$. Using computing tools to solve the one-dimensional minimization in (A30), the global minimum is achieved at $h = \underline{h}$ and $x = \underline{h} + 0.5 - \delta$, where $\delta > 0$ is infinitesimal. The existence of an infinitesimal δ is to enforce the relation $h = \lfloor x + 0.5 \rfloor$. However, for obtaining the minimum value, which will just serve a bound in our purpose, one essentially can just plug $(h, x) = (\underline{h}, \underline{h} + 0.5)$ into (A29) to find the value of the minimum.

$$(v) \quad R_5^{\max} = \tilde{p}_{2u} + \tilde{p}_{1d}$$

$$\begin{aligned} & \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(R_5^{\max}(\theta_1, \theta_2) \right) \\ &= \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(\frac{1}{2} \left(\frac{x^2}{c_1^2 h^2} + \frac{x^2}{c_2^2 h^2} \right) + \frac{1}{2} \left(\frac{\epsilon_2}{h} + \frac{1}{2} \right)^2 - \frac{1}{8} \right) \end{aligned} \tag{A31}$$

$$= \frac{1}{2} \left(\frac{1}{c_1^2} + \frac{1}{c_2^2} \right) + \min_{x,h} \frac{h^2}{x^2} \left(\frac{1}{2} \left(\frac{-1}{2h} + \frac{1}{2} \right)^2 - \frac{1}{8} \right) \tag{A32}$$

$$= \frac{1}{2} \left(\frac{1}{c_1^2} + \frac{1}{c_2^2} \right) + \frac{\hat{h}^2}{(\hat{h} - 0.5)^2} \left(\frac{-\hat{h} - 0.5}{4\hat{h}^2} \right),$$

where, in the minimization of (A31) over ϵ_2 , the minimum is achieved at $\epsilon_2 = -0.5$. Following the steps described in (iii) to minimize (A32) as a one-dimensional problem using computing tools, the global minimum is achieved at $(h, x) = (\hat{h}, \hat{h} - 0.5)$, where $\hat{h} = \max(h, 2)$. This term corresponds to w_3^h . It can be seen that w_3^h is smaller than the term in (A28). Thus, R_3^{\max} does not contribute to $\bar{\rho}^h$.

$$(vi) \quad R_6^{\max} = 1 - (\tilde{p}_{2d} - \tilde{p}_{1d}) - (\tilde{p}_{1u} - \tilde{p}_{2u}).$$

$$\min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(R_6^{\max}(\theta_1, \theta_2) \right) = \min_{x,h,\epsilon_2} \frac{h^2}{x^2} \left(1 + \frac{\epsilon_2}{h} \right) \tag{A33}$$

$$= \min_{x,h} \frac{h^2}{x^2} \left(1 - \frac{1}{2h} \right) = \frac{(h)^2}{(h + 0.5)^2} \left(1 - \frac{1}{2h} \right) \tag{A34}$$

where, in the minimization in (A33) over ϵ_2 , the minimum is achieved at $\epsilon_2 = -0.5$. Using computing tools to solve the one-dimensional minimization in (A34), the global minimum is achieved at $h = \underline{h}$ and $x = \underline{h} + 0.5 - \delta$, where $\delta > 0$ is infinitesimal. Like in (iv), one can simply plug $(h, x) = (\underline{h}, \underline{h} + 0.5)$ into (A34). This corresponds to w_4^h , which is always smaller than the term in (A30). Thus, R_4^{\max} does not contribute to $\bar{\rho}^h$.

References

- Akyildirim, Erdinç, Yan Dolinsky, and H. Mete Soner. 2014. Approximating stochastic volatility by recombining trees. *The Annals of Applied Probability* 24: 2176–205. [CrossRef]
- Andersen, Leif B. G. 2007. Efficient Simulation of the Heston Stochastic Volatility Model. Available online: <http://ssrn.com/abstract=946405> (accessed on 21 May 2021).
- Ball, Clifford, and Antonio Roma. 1994. Stochastic volatility option pricing. *Journal of Financial and Quantitative Analysis* 29: 589–607. [CrossRef]
- Beliaeva, Natalia A., and Sanjay K. Nawalkha. 2010. A simple approach to price American options under the Heston stochastic volatility model. *Journal of Derivatives* 17: 25–43. [CrossRef]
- Boyle, Phelim P. 1986. Option valuation using a three-jump process. *International Options Journal* 3: 7–12.
- Boyle, Phelim P. 1988. A lattice framework for option pricing with two state variables. *Journal of Financial and Quantitative Analysis* 23: 1–12. [CrossRef]
- Broadie, Mark, and Özgür Kaya. 2006. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* 54: 217–31. [CrossRef]
- Burgin, Mark, and Gunter Meissner. 2012. Negative Probabilities in Financial Modeling. *Wilmott Magazine* 58: 60–65.
- Chung, San-Lin, and Pai-Ta Shih. 2007. Generalized Cox-Ross-Rubinstein binomial models. *Management Science* 53: 508–20. [CrossRef]
- Costabile, Massimo, Ivar Massabø, and Emilio Russo. 2012. A forward shooting grid method for option pricing with stochastic volatility. *Journal of Derivatives* 20: 67–78. [CrossRef]
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53: 385–408. [CrossRef]
- Cox, John C., Stephen Ross, and Mark Rubinstein. 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7: 229–64. [CrossRef]
- Durrett, Richard. 1996. *Stochastic Calculus: A Practical Introduction*. Boca Raton: CRC Press Inc.
- Feller, William. 1951. Two singular diffusion problems. *Annals of Mathematics* 54: 173–82. [CrossRef]

- Haug, Espen Gaarder. 2007. Why so negative to negative probabilities. In *Derivatives Models on Models*. New York: John Wiley & Sons, Chapter 14.
- Heston, Steven L. 1993. A closed-form solutions for options with stochastic volatility with application to bond and currency options. *The Review of Financial Studies* 6: 327–43. [CrossRef]
- Hull, John C., and Alan White. 1988. The use of control variate technique in option-pricing. *Journal of Financial and Quantitative Analysis* 23: 237–51. [CrossRef]
- Hull, John C., and Alan White. 1990. Valuing derivative securities using the explicit finite difference method. *Journal of Financial and Quantitative Analysis* 25: 87–100. [CrossRef]
- Hull, John C., and Alan White. 1993. One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis* 28: 235–54. [CrossRef]
- Hull, John C., and Alan White. 1994. Numerical procedures for implementing term structure models II: Two-factor models. *Journal of Derivatives* 2: 37–49. [CrossRef]
- Kouritzin, Michael A. 2000. Exact infinite dimensional filters and explicit solutions. In *Stochastic Models*. Edited by Luis G. Gorostiza and B. Gail Ivanoff. Providence: American Mathematical Society, pp. 265–82.
- Kouritzin, Michael A., and Anne Mackay. 2020. Branching particle pricers with Heston examples. *International Journal of Theoretical and Applied Finance* 23: 2050003. [CrossRef]
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies* 14: 113–47. [CrossRef]
- Maghsoodi, Yoosef. 1996. Solutions of the extended CIR term structure and bond option valuation. *Mathematical Finance* 6: 89–109. [CrossRef]
- Rendleman, Richard J., and Brit J. Bartter. 1979. Two-state option pricing. *Journal of Finance* 34: 1093–110. [CrossRef]
- Rough, Fabrice D. 2013. *The Heston Model and Its Extension in Matlab and C#*. Hoboken: Wiley.
- Ruckdeschel, Peter, Tilman Sayer, and Alexander Szimayer. 2013. Pricing American options in the Heston model: A close look on incorporating correlation. *Journal of Derivatives* 20: 9–29. [CrossRef]
- Tseng, Chung-Li, and Kyle Lin. 2007. A framework using two-factor price lattices for generation asset valuation. *Operations Research* 55: 234–51. [CrossRef]
- Zvan, R., Peter A. Forsyth, and K. R. Vetzal. 2001. Negative Coefficients in Two Factor Option Pricing Models. Working Paper. Available online: <https://cs.uwaterloo.ca/~paforsyt/posmesh3.pdf> (accessed on 21 May 2021).