Valuing American options using fast recursive projections

Cosma, Antonio; Galluccio, Stefano; Pederzoli, Paola; Scaillet, Olivier

Valuing American options using fast recursive projections

Antonio Cosma[*], Stefano Galluccio[**],
Paola Pederzoli[***], and Olivier Scaillet[***]

[*]University of Luxembourg
[**]Incipit Capital, London
[***]University of Geneva and Swiss Finance Institute

March 2, 2016

Abstract

We introduce a fast and widely applicable numerical pricing method that uses recursive projections. We characterize its convergence speed. We find that the early exercise boundary of an American call option on a discrete dividend paying stock is higher under the Merton and Heston models than under the Black-Scholes model, as opposed to the continuous dividend case. A large database of call options on stocks with quarterly dividends shows that adding stochastic volatility and jumps to the Black-Scholes benchmark reduces the amount foregone by call holders failing to optimally exercise by 25%. Transaction fees cannot fully explain the suboptimal behavior.

Keywords : Option pricing, American option, Bermudan option, discrete transform, discrete dividend paying stock, suboptimal non-exercise, numerical techniques.
JEL classification: G13, C63.

We thank J. Detemple, D. Duffie, E. Gobet, F. Moraux, D. Newton, A. Treccani and A. Valdesogo for valuable insight and help in addition to the participants at the 6th World Congress of the Bachelier Finance Society in Toronto, the 2011 European Econometric Society in Oslo, the 2013 Mathematical Finance Day in Montreal, the 2014 Mathematical and Statistical Methods for Actuarial Sciences and Finance in Salerno, the 2014 International Symposium on Differential Equations and Stochastic Analysis in Mathematical Finance in Sanya, the 7th General Advanced Mathematical Methods in Finance and Swissquote Conference in Lausanne, the 2015 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics in Cape Town, the 9th International Conference on Computational and Financial Econometrics in London, the 2015 International Conference on Computational Finance in London, and seminars at the University of Geneva and the University of Orléans. O. Scaillet received support from the Swiss NSF through the NCCR Finrisk. Paola Pederzoli acknowledges the financial support of the Swiss NSF (grant 100018 − 149307).

This paper provides financial economists and practitioners with a new, fast and convergent numerical method to solve a wide range of problems in derivative pricing. The derivatives we are able to price are contracts whose path is monitored at discrete moments in time. This class is by no means restrictive. As outlined in Andricopoulos, Widdicks, Newton, and Duck (2007), few pricing problems of practical interest fall outside the scope of the application of our method. The challenges in studying derivative products come from the following two sources: the degree of sophistication of the process of the underlying asset and the complexity of the contracts in terms of payoff and exercise rules. Almost any departure from the plain vanilla European style options implies that closed-form pricing formulas are no longer available (for an extensive review, see Detemple (2005)). Most works introducing new numerical methods claim universality and speed but typically achieve them in only one of the two dimensions. There are a few works, such as Longstaff and Schwartz (2001), Haugh and Kogan (2004), and Chen, Härkönen, and Newton (2014), that tackle the two dimensions of complexity in a highly competent manner. However, they seldom make it to the application stage, and when they do so, it is more often in the quantitative desks of financial institutions rather than in empirical academic research. Here, we wish to provide colleagues and practitioners with a reliable empirical tool. Applicability has been the main goal in its development. Our method, which we refer to as recursive projections, is faster than the existing methods and can accommodate both sophisticated dynamics for the underlying asset and path-dependent payoffs. To show the full potential of our technology, we extend the empirical work of Pool, Stoll, and Whaley (2008) on the observed suboptimal non-exercise of American call options written on dividend-paying stocks. Far from being merely an illustrative example, the empirical part of our paper is a scientific contribution in its own right. We show that, by taking into account stochastic volatility and jumps in the process of the underlying asset, we can explain up to 25% of the gain forgone due to non-optimal exercise decisions, as computed in Pool, Stoll, and Whaley (2008). Because financial frictions are a possible explanation of departure from the expected exercise behavior (e.g. Jensen and Pedersen (2014)), we also show that transaction costs cannot fully explain the non-exercise decisions. In the process, to our knowledge, we are the first to provide comprehensive descriptive statistics of the parameters driving the jumps and the stochastic volatility of the constituents of the Dow Jones Industrial Average Index (DJIA) traded in the period from January 1996 to December 2012. To date, the same parameters have been estimated only for equity indexes (see Bakshi, Cao, and Chen (1997); Eraker, Johannes, and Polson (2003); Broadie, Chernov, and Johannes (2009)), because the computations are easier than those for single stocks. Thus, our findings in themselves are a valuable reference for future empirical research on dividend-paying stocks. We also provide new theoretical insights into how the early exercise boundary changes, depending on the discrete or continuous nature of the dividend distributed by the underlying asset. In particular, we show that the boundary

is higher under the Merton (1976) jump-diffusion and Heston (1993) stochastic volatility models than under the Black-Scholes model if the dividend is discrete, whereas we know it is lower in the case of a continuous dividend yield (Amin (1993); Adolfsson, Chiarella, Ziogas, and Ziveyi (2013)). The study of the early exercise boundary is important for investment decisions. For example, Battauz, De Donno, and Sbuelz (2014) characterize the double continuation region implied by an option with a negative interest rate, which occurs in the case of gold loans.

We can now quantify our claim concerning the speed of recursive projections. We need only two matrix-per-vector multiplications and a couple of seconds to evaluate an American call option and its Greeks on a stock paying one dividend before the expiry date in the Black-Scholes case. The competing (non-recombining) tree method requires thousands of steps and hundreds of seconds to yield the same accuracy. Similarly, a few seconds are sufficient to evaluate an American call option and its Greeks on a stock paying three dividends before the expiry date in the Heston model, whereas we would need thousands of steps and more than a hundred seconds to yield the same accuracy using the finite-differences method alternative. To date, no empirical work on options written on dividend-paying stocks exists outside the Black-Scholes world, given that the existing methods are simply too time-consuming.

To obtain a first intuition on how our method works, we can view the pricing of a derivative security essentially from two perspectives, with the link between the two being given by the Feymann-Kac theorem. The first perspective is solving the partial differential equation (PDE) to yield the price of the derivative assets. Numerically discretizing the differential operator leads to finite difference schemes (see Brennan and Schwartz (1977); Clarke and Parrott (1999); Ikonen and Toivanen (2008)). This method is the most common approach in regard to numerically finding solutions to complex pricing problems, and we benchmark our method against the PDE outside the Black-Scholes setting. The second perspective is viewing the price of the derivative asset as the conditional expectation of the discounted future payoff. It exploits the knowledge of the discounted probability distribution (the so-called Green function) with respect to which the conditional expectation is taken. The advantage of this class of methods over the previous class is that we only need to evaluate the value function at specific points in times, typically potential exercise dates, and not repeatedly at infinitesimally small consecutive time steps. This paper adopts the second perspective. We share some similarities with quadrature-based methods that, in recent years, have provided effective routines to price path-dependent options. Newton and co-authors follow up on an early intuition by Sullivan (2000) and provide a pricing routine, called QUAD, in the Black-Scholes framework (Andricopoulos, Widdicks, Duck, and Newton (2003); Andricopoulos et al. (2007)), and eventually extending it to a wider range of modelling environments for the underlying (Chen et al. (2014)). Before the latter contribution, O'Sullivan (2005) and Lord, Fang,

Bervoets, and Oosterlee (2008) made two early attempts at extending the quadrature framework beyond Black-Scholes. Their elegant efforts are somewhat limited in scope, being unable for instance to price path-dependent options in the stochastic volatility framework, a very common requirement in empirical research.

We cast the problem of the computation of the conditional expectation in a functional projection framework. Doing so permits the representation of the conditional expectation operator in the simplest form of standard linear operators, i.e., matrices, and the pricing of derivative contracts by means of linear algebra tools. Functional projection on localized basis functions occurs through the simple sampling of the relevant functions at given grid points. Although, in its simplest form, our method resembles quadratures, it is more general in making possible to derive a simple recursive formula to price path-dependents contracts. Indeed, we show that the appealing feature of our method is its recursive nature; that is, pricing boils down to a sequence of matrix time vector multiplications. The output of a multiplication is the input of the following step, without the need for intermediate computations, such as the careful placing of knots at discontinuities in the quadrature. The placing of intermediate knots in quadratures can be cumbersome if payoffs include dividends, as in our leading application. This lack of intermediate computational overhead boosts the speed of our method and makes it feasible for empirics. We can accommodate all modelling choices for the underlying, provided that the transition density is analytically known, either in the direct space or in some transform space (Fourier or Laplace). Fourier transforms of transition probabilities describe price evolution in affine models (Duffie, Pan, and Singleton (2000)), quadratic models (Leippold and Wu (2002); Cheng and Scaillet (2007)), and variance gamma and Levy models (Madan, Carr, and Chang (1998); Carr, Geman, Madan, and Yor (2003)). In addition, projecting transition densities and value functions on appropriate function bases, allows us to naturally describe the filtration of information at given points in time, thus correctly addressing non-Markovian models, such as stochastic volatility models, in the stock price dimension. This feature is the hurdle O'Sullivan (2005) and Lord et al. (2008) could not cross. Moreover, our projection method makes the numerical error in approximating the transition densities negligible for any time horizon. We can evaluate transitions of the value functions for arbitrary time steps, whereas Chen et al. (2014) can at most address time steps equal to 0.1 year. This pricing approach adds intermediate steps that slow down their pricing algorithm. To conclude, our contribution is a novel and general stochastic optimum control method that contains, as particular cases, most of the advances in numerical option pricing based on quadratures. Moreover, its matrix form has an appealing interpretation in terms of the stochastic discount factor, which means that the applied researcher has not only a simple and fast tool at hand but also an intuitive understanding of the object he is using.

The paper is organized as follows. The following section reviews the most recent ad-

vances in numerical option pricing. In Section II.C, we develop an introductory example based on the Black-Scholes model and present some preliminary numerical results that show the advantages of our technique. In Section III, we study the general case of valuation by fast recursive projections in order to include standard affine models. We present numerical illustrations for the Black-Scholes and Heston models. We also provide the theoretical convergence of the computed option price in terms of the sampling frequency, and characterize the convergence rate of the computed option price. In Section IV, we present the innovative applications of our algorithm. In Section IV.A, we characterize the early exercise boundary under different modelling assumptions. In Section IV.B, we present the empirical results concerning the cost of failing to optimally exercise American call options. Section V gathers some concluding remarks. The supplementary online Appendix contains the proofs of the propositions, additional comparisons with existing methods and gives a detailed description of the data and the calibration procedure.

I. Review of alternative pricing methods

As explained in the introductory section above, the numerical methods for option pricing fall into the following two categories: differential methods and integral methods. Differential methods provide the solution to the pricing problem by numerically approximating the associated partial differential equation (see the references of the previous section). Integral methods approximate the conditional expectation, giving the arbitrage-free value of a financial contract. Some methods lie at the boundary of the two groups. The binomial tree technology, and lattice methods in general (Cox, Ross, and Rubinstein (1979); Broadie and Detemple (1996); Vellekoop and Nieuwenhuis (2006)), although similar to finite differences in their implementation, belong to the integral family of methods, given that they provide a discrete approximation of a conditional expectation. Barone-Adesi and Whaley (1987) start from the partial differential equation representation in the Black-Scholes case, and contribute a closed form approximation for an American option with continuous dividend yield.

In this section, we focus on integral methods because the main recent contributions, including ours, fall within this family. The integral representation arises from the observation that option valuation in arbitrage-free economies amounts to using linear operators that assign prices today to payoffs at futures dates. These linear operators are conditional expectation operators. In multi-period economies, the temporal consistency in valuation ensures that the family of such operators satisfies a semigroup property (Garman (1985)). Because most valuation models correspond to Markov environments for the price of the underlying assets, the semigroup property is formalized by the law of iterated expectations restricted to Markov processes (Hansen and Scheinkman (2009)), where the Markov states are indexed by the time horizons of potential exercise. From a

computational perspective, pricing algorithms attempt to approximate the following: (i) the conditional expectation with respect to the information at one potential exercise date and (ii) the recursive relationship between conditional expectations at different potential exercise dates. Geske and Johnson (1984) express the problem of pricing an option with discrete exercise dates as a multivariate integral. Kim (1990), Jamshidian (1992), and Carr, Jarrow, and Myneni (1992) extend this approach to continuous time and express the early exercise premium of American contracts as an integral term whose interpretation is the risk-neutral valuation of a continuous cash flow. This additional integral is a function of the exercise boundary. Bunch and Johnson (2000), Huang, Subrahmanyam, and Yu (1996), Ju (1998), and Ibáñez (2003) contribute algorithms to compute the free boundary and the option prices in the Black-Scholes model environment. Detemple and Tian (2002) extend these contributions to diffusion processes with stochastic interest rates. These approaches cannot easily accommodate discrete dividends. Roll (1977), Geske (1979), and Whaley (1981) provide a closed-form approximation of pricing an American option paying a single discrete dividend in the Black-Scholes framework. Medvedev and Scaillet (2010) develop an approximation of American option prices under stochastic volatility and stochastic interest rates using short-term asymptotics (for the Black-Scholes case, see Lamberton and Villeneuve (2003)). These approximations are accurate and flexible, but cannot accommodate discrete dividends. They are mainly suited for index and exchange rate options.

Another approach in approximating the pricing operator is simulating the trajectories of the underlying asset to compute the conditional expectation. One of the first simulation-based analyses of American options in the Black-Scholes case is that of Broadie and Glasserman (1997). A recent stream of Monte Carlo methods started with the work of Longstaff and Schwartz (2001), developing into the so-called duality approach of Haugh and Kogan (2004), Rogers (2002), and Andersen and Broadie (2004). The duality approach method uses simulations to find a lower bound and an upper bound of the true price. In this sense, it is an extension of Broadie and Detemple (1996). As shown in the numerical experiments of Andersen and Broadie (2004), the interval can be tight, which makes the method very precise. The advantage of this approach is that it can handle virtually any type of process dynamics, state variable structure, and payoff specification. However, to achieve precision, as in all simulation-based methods, the discretization step in the time dimension must be very small, making computations lengthy. Desai, Farias, and Moallemi (2012) refine the duality approach by contributing a new algorithm to compute the upper bound, improving it in terms of speed.

We can express conditional expectations in terms of series developments on appropriate basis functions, which is the approach we follow when projecting the Green and value functions on localized basis functions in the subsequent sections. We can already find simple and tractable projections of complex payoffs in the literature via the repre-

sentation on a set of basis functions (such as polynomials in Madan and Milne (1994); Lacoste (1996); Darolles and Laurent (2000)). Chiarella, El-Hassan, and Kucera (1999) suggest a fast recursion projection method based on Hermite polynomials in the Black-Scholes model (for an extension of the model in Merton (1976), see Chiarella and Ziogas (2005)). They also explain how their method provides a viable numerical method for the implementation of the path integral approach to option pricing as described in Linetsky (1997), for example. Our paper is an extension of the former in that we use sampling instead of projections to improve speed. This family of models typically assumes a geometric Brownian motion for the underlying asset, with some exceptions accommodating for jumps but not for stochastic volatility.

In a more recent application of Hermite polynomials to pricing, Xiu (2014) provides a general, closed-form approximation of European, but not path-dependent, option prices by using a finite-term expansion of the transition density (for related work on expansions, see also Aït-Sahalia and Kimmel (2007, 2010)). Kristensen and Mele (2011) approximate the option price by expanding the difference between the true model price and the Black-Scholes price. They can price some specific path-dependent options, such as barrier options but cannot address more general path-dependent payoff features, such as discrete dividends. Our recursion projection method shares some features with the dynamic programming approach of Ben-Ameur, Breton, and Martinez (2009) in that they also approximate the value function on a fixed grid and interpolate with local polynomials to reconstruct the value surface. They assume a GARCH process for the stochastic volatility and obtain closed-form formulas for the conditional expectations to compute the grid of values at a previous time. Though computationally efficient, their method is restricted to the GARCH family of processes and does not naturally accommodate dividends.

As explained above, our method shares features with the quadrature family of models. In addition to the contributions cited above, Jackson, Jaimungal, and Surkov (2008) introduce a Fourier space time stepping algorithm to price Bermudan options under Levy processes. This method shares advantages and shortcomings with the method of Lord et al. (2008), in that it is unable to price path-dependent options when the underlying asset follows a stochastic volatility process.

II. An introductory example

In this section, we show how a pricing problem for a European payoff in the Black-Scholes model translates into a functional projection. The derivation uses elementary calculus concepts. Then, we explain how we exploit the projection to build fast recursive schemes to value path-dependent options such as Bermudan and American options when the stock pays discrete dividends. We design this introductory example to emphasize the intuition underpinning our approach.

A. *Description of the method: the European case*

Let S_t be the price of the underlying asset at date t and assume that interest rates are constant to facilitate exposition. For $S_t = x$, the value function $V(x,t)$ for a European option is given by the following conditional expectation:

$$V(x,t) = \mathbb{E}\big[e^{-r(T-t)}H(S_T,T)\big|S_t = x\big], \tag{1}$$

where $H(S_T,T)$ is the payoff function expressed as a function of time T and of the value of the underlying asset S_T at maturity date T, and r is the constant risk-free interest rate. When the pricing operator in (1) admits a state price density $G(x,t;y,T)$, the so-called Green function, which is the discounted value of the transition probability density from point x at time t to point y at time T, we obtain the familiar integral form:

$$V(x,t) = \int G(x,t;y,T)H(y,T)dy. \tag{2}$$

Now consider a regularly spaced grid of points $\{y_1, y_2, \ldots, y_N\}$ that defines a finite interval $D = [y_1 - \Delta y/2, y_N + \Delta y/2]$, with $\Delta y = y_j - y_{j-1}$. We know that, under appropriate regularity conditions, the integral (2) restricted to the interval D can be approximated by the Riemann sum as follows:

$$V(x,t) \sim \sum_{j=1}^{N-1} G(x,t;y_j,T)\, H(y_j,T)\, \Delta y, \qquad t < T, \tag{3}$$

where the $\sim$ symbol means that the right hand term converges to the left hand term as $\Delta y \to 0$. The representation (3) is known as the 'rectangle method' in standard integral calculus. If we are interested in computing the value of $V(x,t)$ on a regularly spaced grid of points $\{x_1, \ldots, x_M\}$, for instance to plot the value function of the contract, or to compute the Greeks, we can express (3) in a matrix form as follows:

$$\mathbf{V}(t) \sim \mathbf{v}(t) = \mathbf{G}(t;T)\mathbf{H}(T), \tag{4}$$

where $\mathbf{V}(t) = \big(V(x_1,t), \ldots, V(x_M,t)\big)'$, $\mathbf{v}(t)$ is the approximation of $\mathbf{V}(t)$ obtained from the $N \times 1$ vector $\mathbf{H}(T)$ with entries $\mathbf{H}_j = H(y_j,T)$ and the $M \times N$ matrix $\mathbf{G}$ with entries $\mathbf{G}_{ij} = \Delta y\, G(x_i,t;y_j,T)$. Equation (4) describes a discretization of the economy. We can think of the matrix $\mathbf{G}$ as a matrix of Arrow-Debreu prices, in which the rows represent discrete states $\{x_i\}_{i=1,\ldots,M}$ at the current date t, and the columns represent discrete states $\{y_j\}_{j=1,\ldots,N}$ at the future date T. We can interpret the column vectors $\mathbf{V}(t)$ and $\mathbf{H}(T)$ as vectors of state contingent values at dates t and T, respectively. The matrix operator $\mathbf{G}$ discounts future payoffs at date T to current prices at date t (Garman (1985)).

Let us more closely examine the elements of $\mathbf{H}(T)$. For readability, we define $\underline{y}_i = y_i - \frac{\Delta y}{2}$ and $\overline{y}_i = y_i + \frac{\Delta y}{2}$. We can interpret the value $H(y_j,T)$ as an approximation of

the quantity

$$(1/\Delta y)\int_{\underline{y}_j}^{\overline{y}_j} H(y,T)dy = (1/\sqrt{\Delta y})\int \frac{1}{\sqrt{\Delta y}}\mathbb{I}_{\underline{y}_j,\overline{y}_j}H(y,T)dy \stackrel{\text{def}}{=} \frac{1}{\sqrt{\Delta y}}\int e_j(y)H(y,T)dy, \tag{5}$$

where $e_j(y) = \frac{1}{\sqrt{\Delta y}}\mathbb{I}_{\underline{y}_j,\overline{y}_j}$, and where $\mathbb{I}_{\underline{y}_j,\overline{y}_j}$ is the indicator function of the interval $[\underline{y}_j, \overline{y}_j)$. Because $\{e_j(y)\}_{j=1,\ldots,N}$ is an orthonormal set given the standard L_2 inner product $\langle f, g\rangle = \int f(x)g(x)dx$, we can view the entries of the vector $\mathbf{H}(T)$ as an approximation of $(1/\sqrt{\Delta y})$ times the coefficients of the decomposition of the payoff function $H(y,T)$ on the set of orthonormal indicator functions defined by the grid $\{y_1,\ldots,y_N\}$. A similar argument can be applied to the coefficients of the $\mathbf{G}(t;T)$ matrix: every row of $\mathbf{G}(t;T)$ is given by an approximation of $\sqrt{\Delta y}$ times the coefficients of the projection of the conditional density $G(x_i,t;y,T)$ on the same orthonormal set. The different choices in the normalization factors for the entries of $\mathbf{G}$ and $\mathbf{H}$ are justified by our wish to interpret all of the quantities appearing in (4) as prices. We have that $V(x,t)$ and $H(y,T)$ are value functions and $G(x,t;y,T)$ is a state price density.

Overall, we can interpret the numerical approximation of the integral (2) as a projection of the functions on an orthonormal basis. Such an interpretation is key in the generalization of the recursive projection approach to more sophisticated models in the next section. In our case, the functional projection boils down to sampling the given functions on a grid of N points $\{y_j\}_{j=1,\ldots,N}$ via a discrete transform. From a computational perspective, the entries of $\mathbf{H}(T)$ summarize how payoffs depend on the price of the underlying assets at future exercise dates, and the entries of $\mathbf{G}(t;T)$ summarize how the price of the underlying assets transits from one state to another according to the elapsed time between exercise dates. Figure 1 gives a graphical representation of the two computational steps of our fast recursive projection approach, (i) the projection step, and (ii) the recursive step. The value function at t, on the right, and the state price density for a given value of $x = S_t$, on the left, are sampled (the projection step); the obtained arrays of values are multiplied element by element, and the products are summed to obtain the value of $V(x,t)$ (the recursive step).

[Figure 1 about here]

B. Description of the method: the path-dependent case

Let us now address the valuation of path-dependent contracts. We start by considering a Bermudan option. We consider a set $\{t_1 = t,\ldots,t_L = T\}$ of exercise dates. At each t_l, the holder of a Bermudan option may decide whether to exercise. He exercises if the intrinsic value $H(S_{t_l},t_l) \stackrel{\text{def}}{=} (S_{t_l}-K)_+ = \max\{S_{t_l}-K,0\}$ is higher than the value of keeping the option, i.e., the continuation value. Bermudan options are the ideal building blocks for studying American call options on dividend-paying stocks. It is well known

that it can be optimal to exercise American call options immediately before ex-dividend dates $\{t_h, t < t_h < T\}_{h=1,\ldots,H}$; for instance, see Pool et al. (2008) for a discussion on early exercise strategies. The implication is that we must monitor the option value function $V(x,t)$ immediately before the ex-dividend dates, when the intrinsic value $(S_{t_h-\epsilon} - K)_+$ for a small $\epsilon > 0$ can be larger than the continuation value $V(S_{t_h}, t_h)$. Then an American call option shares with a Bermudan option the feature that its value function must be evaluated only at a finite number of dates.

The semigroup property of the pricing operator ensures that we compute the value function $V(x,t)$ of a Bermudan option recursively. The recursion consists of moving backwards in time and computing at each $t_l, l = 1, \ldots, L-1$:

$$V(x, t_l) = \max\Big\{H(x, t_l), \mathbb{E}\big[e^{-r(t_{l+1}-t_l)}V(S_{t_{l+1}}, t_{l+1})|S_{t_l} = x\big]\Big\}, \tag{6}$$

with the boundary condition $V(y, t_L) = H(y, t_L)$. To speed up the recursion, we impose the condition that the grid of values $\{y_j\}_{j=1,\ldots,N}$ at which we sample the function $V(y, t_{l+1})$ and the grid $\{x_i\}_{i=1,\ldots,M}$ at which we compute the function $V(x, t_l)$, coincide at each exercise date, which means that $M = N$. From now on, we use the x variable as a generic conditioning value, e. g. the value of the underlying at date t, as in $V(x,t)$. If x takes a specific value in the grid $\{y_i\}_{i=1,\ldots,N}$, then we write $V(y_i, t)$. Then, in the matrix notation of the approximation, we obtain the following:

$$\mathbf{V}(t_l) \sim \mathbf{v}(t_l) = \max\Big\{\mathbf{H}(t_l), \mathbf{G}(t_l; t_{l+1})\mathbf{v}(t_{l+1})\Big\}, \tag{7}$$

and the approximation $\mathbf{v}(t_l)$ is the input for the following recursion step to compute an approximation $\mathbf{v}(t_{l-1})$ of $\mathbf{V}(t_{l-1})$. Figure 2 shows how combining projection and recursion steps with identical grids at each exercise date translates the pricing problem of path-dependent options into a sequence of matrix time vector multiplications.

[Figure 2 about here]

The convergence properties of $\mathbf{v}(t_l)$ to $\mathbf{V}(t_l)$ are formally established in Proposition 1 of Section III. From (7), it is clear that computations only occur at the exercise dates defined in the Bermudan contract, and do not require any input at any other point in time. Furthermore, under the previous additional constraints on the grids, if the time interval $\tau = t_{l+1} - t_l, l = 1, \ldots, L-1$, is constant and if the pricing operator enjoys a stationarity property (time translation invariance), then the matrix $\mathbf{G}(t_l; t_{l+1}) = \mathbf{G}(\tau)$ has constant entries, and the algorithm only involves one single computation of the matrix.

The methodology easily extends in the presence of discrete dividends paid on potential exercise dates[1]. The implication is that the set of ex-dividend dates is a subset of the Bermudan exercise dates and that we have $\{t_h\}_{h=1,\ldots,H} \subset \{t_l\}_{l=1,\ldots,L}$. We only need

[1]We can easily extend to the case in which the ex-dividend date does not belong to the set of exercise dates. We do not explicitly state this case, given that we are mainly interested in Bermudan options, as a building block for studying American options.

to add the dividend δ to the continuation value in (6). Hence, in order to price an American option on a dividend-paying stock, Equation (7) must be modified by sampling the state price density $G(x, t_l; y, t_{l+1})$ at the grid $\{y_i - \delta(y_i)\}_{i=1,\dots,N}$ for the conditioning value x, whenever $t_l \in \{t_h\}_{h=1,\dots,H}$. The entries of the matrix $\mathbf{G}(t_l; t_{l+1})$ then become $\mathbf{G}_{ij} = G(y_i - \delta, t_l; y_j, t_{l+1})\Delta y$. Given the freedom in choosing where to sample G, $\delta(x)$ could be any function of x. If $\delta(x) = r_d\, x$, then we can accommodate for a proportional dividend. If $\delta(x) = d$, then we can accommodate for a discrete dividend amount d. If $\delta(x) = 0$, then we return to the Bermudan option case. The value function $\mathbf{V}(t_h)$ still gives the value of the contract at the grid points $\{y_1, \dots, y_N\}$; thus we can use its approximation $\mathbf{v}(t_h)$ as the input for the following step of the algorithm, and the recursive property of the algorithm is maintained. Figure 3, shows how the recursive scheme changes to accommodate for dividends.

[Figure 3 about here]

The early exercise decision for American long put holders is more complicated. We consider a Bermudan put with exercise dates $\{t_1, \dots, t_L\}$, and assume that the dates $\{t_h\}$ at which dividends are paid form a subset of the exercise dates. Then, by taking L large, our approach also provides a quick approximation for pricing American put options on single stocks paying discrete dividends. In the Black-Scholes model, the state price density is known in the following closed form:

$$G(x, t; y, T) = \frac{1}{y} \frac{e^{-r(T-t)}}{\sqrt{2\pi\sigma^2(T-t)}} \exp\left(-\frac{\big(\log y - \log x - (r - \sigma^2/2)(T-t)\big)^2}{2\sigma^2(T-t)} \right), \qquad (8)$$

where σ is the volatility and the pricing operator satisfies the stationarity property. Hence, we expect to obtain a fast, simple and accurate numerical algorithm. Before reporting the numerical results, we stress that we have kept the setup of this introductory example as simple as possible to emphasize intuition by limiting technical details. Although the computational results already speak in favor of our construction, it is in the following sections that the recursive projection method deploys its full potentiality with more complex price dynamics.

C. Numerical illustrations in the Black-Scholes model

As a first numerical example in the Black-Scholes framework, we compare the convergence speed of a binomial tree and of the recursive projection method in pricing an American call option on a dividend-paying stock[2]. Two popular methods are a known cash amount d or a known dividend yield r_d. The latter is computationally friendly because it leads to a recombining tree. However, the empirical evidence shows that cor-

[2]All of the codes are written in C++. The codes are available from the authors upon request. Codes for the binomial tree are taken from Ødegaard (2014).

porations tend to commit to paying out fixed amounts at regular dates and to smooth their dividends rather than adjusting them downwards and signaling a decrease in cash flows (for a signaling based theory on dividend policy see, for instance, Miller and Rock (1985)). These arguments justify our preference for modeling dividends as fixed known amounts rather than as given yields. The known dividend amount assumption does not lead to a recombining tree, and a new tree is originated at each node following an ex-dividend date, increasing the numerical complexity of the problem. If ν is the number of time discretization steps, the number of nodes of a recombining tree grows as ν^2, and the number of nodes of a non recombining tree grows as ν^{H+2}, where H is the number of ex-dividend dates.

However, the number of operations in the recursive projection method grows linearly with the number H of dividends paid because every dividend date adds only a single extra step in the recursive algorithm.

[Figure 4 about here]

Figure 4 compares the convergence speed of a binomial tree and that of the recursive projection method in pricing an American call option on a discrete dividend paying stock. The option has a maturity of $T = 3$ years and a dividend $d = 2$ is paid out at the end of each year. Other parameters, namely the interest rate, volatility and strike price, are set equal to $r = 0.05$, $\sigma = 0.2$, and $K = 100$, respectively. We compute 3 prices: at-the-money, in-the-money and out-of-the-money, corresponding to $S_0 = 80, 100$, and 120, respectively. The *true* values of 7.180, 18.526, and 34.033 are obtained with 10000 time steps in the binomial tree. The graphs show that, across the three different values of S_0, the recursive projections enjoy a much faster speed with a factor of approximately 10^4 for a comparable level of precision. The speed advantage is even larger if we consider that a new tree is needed for each value of S_0. Instead, the recursive projection method delivers the entire value function $\mathbf{v}(0)$ at once in a straightforward manner. This feature is particularly useful in computing Greeks through numerical differentiation. As we have argued, the advantage of recursive projections over binomial trees is that the former method can address several dividend payment dates, whereas we can reasonably use the latter for, at most, two dividend payment dates. In Section E of the supplementary online Appendix, we extend the above pricing exercise, and further compare the recursive projections with two more recent methods, namely the modified binomial tree proposed by Vellekoop and Nieuwenhuis (2006) and the Monte Carlo based method by Longstaff and Schwartz (2001). With respect to the modified binomial tree, the recursive projections are still faster, even if the results are less striking than in the case of the non-recombining tree. The recursive projections are one order of magnitude faster; nevertheless, this difference is somewhat underestimated because the Vellekoop and Nieuwenhuis method shares the shortcomings of the non-recombining tree, namely, that (i) a new tree needs to

be computed for each S_0 and (ii) it cannot be easily extended beyond the Black-Scholes case. The recursive projections are faster than the Longstaff Schwartz method by at least four orders of magnitude. As an aside, if for $S_0 = 100$, we approximate the known constant dividend $d = 2$ with a known continuous dividend yield[3] $r_d = 0.013$, then a binomial tree with 10000 steps delivers a value of 18.213 instead of 18.526, with a relative error of approximately 169*bp*. This error is far above observed bid-ask spreads. This simple example points to the importance of using models that can explicitly address discrete dividends in empirical analysis, instead of using approximations based on continuous dividend yields. In Section IV, we more extensively analyze the impact on the exercise boundary of the choice of a continuous dividend yield or a discrete dividend.

Moreover, we have chosen a sampling scheme that is equivalent to projecting the payoff function on a set of basis functions that are well localized, in the sense that their support is a closed interval. The implication is that local features of the payoff function, such as a discontinuity, are described by the coefficients relative to one or at most two basis functions, those lying next to the discontinuity. This description avoids a noisy approximation induced by spurious oscillations when projecting discontinuities on basis functions defined on the entire domain, such as the Fourier sine-cosine basis or the Hermite polynomial basis. From a computational perspective, this property translates into an accurate approximation even for payoffs with strong discontinuities, such as a digital payoff $H(S_{t_l}, t_l) = \mathbb{I}_{S_{t_l} > K}$ in a Bermudan digital call option. Figure 5 (see the caption of the table for the values of the parameters of the example) shows that the binomial tree has problems capturing the discontinuity in the payoff function. Consequently, an extremely slow convergence of the tree method for at-the-money Bermudan digital call options is yielded. The recursive projections are also at least an order of magnitude faster in pricing the out-of-the-money options. The apparent non-monotonic convergence of the binomial tree for $S_0 = 120$ is because both methods achieve a quick convergence for in-the-money options, and the graph only displays small oscillations on the order of half a basis point around the true value.

[Figure 5 about here]

III. Valuation by fast recursive projections

In this section, we generalize the approach developed in the introductory example of Section II.A. We consider a set of exercise dates $\{t_l\}_{l=1,\ldots,L}$, with $t_L = T$. The starting

[3]The yield is obtained by considering the dividends paid at $t = 1$ and $t = 2$ only, because the dividend paid at $t = 3$ has no impact on the price of the option. Considering a dividend yield of 2% would provide an option value of 16.857, which is a much larger error.

point is the following identity:

$$V(x,t) = \int G(x,t;y,T)H(y,T)dy = \int \Big(\sum_{i=-\infty}^{\infty} \langle G, \varphi_i \rangle \varphi_i(y) \Big) \Big(\sum_{j=-\infty}^{\infty} \langle H, \varphi_j \rangle \varphi_j(y) \Big) dy$$

$$= \sum_{i,j=-\infty}^{\infty} \langle G, \varphi_i \rangle \langle H, \varphi_j \rangle \int \varphi_i(y)\varphi_j(y)dy = \sum_{j=-\infty}^{\infty} \langle G, \varphi_j \rangle \langle H, \varphi_j \rangle, \tag{9}$$

where $\{\varphi_j\}_{i\in\mathbb{Z}}$ is a generic orthonormal basis in L_2, and $\langle G(x,t;\cdot,T), \varphi_j(\cdot)\rangle$ and $\langle H(\cdot,T), \varphi_j(\cdot)\rangle$ (abbreviated in $\langle G, \varphi_j\rangle$ and $\langle H, \phi_j\rangle$ respectively) are the coefficients of the linear projection of G and H on $\{\varphi_j\}_{i\in\mathbb{Z}}$. The valuation by fast recursive projections is based on two steps: a projection step and a recursive step. The projection step consists of projecting G and H at time T on the basis $\{\varphi_j\}_{j\in\mathbb{Z}}$ and computing the coefficients $\langle G, \varphi_j\rangle$ and $\langle H, \varphi_j\rangle$. The recursive step consists of transmitting the coefficients back in time by computing the final sum in Equation (9). We develop the method by considering the orthonormal set of indicator functions $\{e_j\}_{j\in\mathbb{Z}}$, $e_j = 1/\sqrt{\Delta y}\, \mathbb{I}_{\underline{y}_j, \overline{y}_j}$ for an equally spaced grid $\{y_i\}$ with step Δy and $\underline{y}_j = y_j - \Delta y/2$, $\overline{y}_j = y_j + \Delta y/2$. The set $\{e_j\}_{j\in\mathbb{Z}}$ is a basis in L_2 when $\Delta y \to 0$. Then, as shown in Equation (5), we can approximate the coefficients $\langle H, e_j\rangle$ and $\langle G, e_j\rangle$ with $\sqrt{\Delta y} H(y_j, T)$ and with $\sqrt{\Delta y}\, G(x,t;y_j,T)$. Thus, the projection step simply consists of sampling the relevant functions at the grid points $\{y_j\}_{j\in\mathbb{Z}}$. The recursion step stems from the observation that, if x in Equation (9) takes values in the grid $\{y_j\}_{j\in\mathbb{Z}}$, then $v(y_i, t_{L-1}) = \sum_{j=-\infty}^{\infty} G(y_i, t_{L-1}; y_j, T)H(y_j, T)\Delta y$ and $v(y_i, t_{L-2}) = \sum_{j=-\infty}^{\infty} G(y_i, t_{L-2}; y_j, t_{L-1})v(y_j, t_{L-1})\Delta y$ are approximations of $V(y_i, t_{L-1})$ and $V(y_i, t_{L-2})$. That is, there is a recursive linear expression that connects the value of the contract at consecutive times and at different points of $\{y_j\}_{j\in\mathbb{Z}}$. Here, all the functions are sampled at the same grid $\{y_j\}_{j\in\mathbb{Z}}$ for each potential exercise date $\{t_l\}_{l=1,\dots,L}$. Doing so allows the matrix representations of Equation (7). In the following sections, we characterize the projection step and the recursive step in two different frameworks. In Sections III.A and III.B, we consider the case in which the state price density is known in closed form, as in the Black-Scholes example developed in Section II or in the Merton jump-diffusion model. In Sections III.C and III.D, we characterize the projection step and the recursion step when the characteristic function of the state price density has a known analytic form, as in the Heston model. In our presentation, we generically refer to this second class of models as to the "stochastic volatility" case. The methodology developed in Sections III.C and III.D covers affine jump-diffusion models and Levy models, in which we price by transform analysis. Here, the key requirement is to be able to numerically compute the characteristic function of the state price density. In Section III.E, we provide numerical examples in the Heston case.

A. *Merton-Black-Scholes: the projection step*

The projection step is based on an approximation of the payoff function and of the state price density by the set of orthonormal functions $\{e_j(y)\}_{j\in\mathbb{Z}}$:

$$\tilde{H}(y,T) = \sqrt{\Delta y} \sum_{j=-\infty}^{\infty} H(y_j,T)e_j(y), \tag{10}$$

$$\tilde{G}(y_i,t;y,T) = \sqrt{\Delta y} \sum_{j=-\infty}^{\infty} G(y_i,t;y_j,T)e_j(y). \tag{11}$$

The values of $\tilde{H}(y,T)$ and $\tilde{G}(y_i,t;y,T)$ in each interval $[\underline{y}_j,\overline{y}_j)$ are given by the values of the functions G and H sampled in the mid point y_j of each interval. Then, we have that $\tilde{H}(y,T)$ and $\tilde{G}(y_i,t;y,T)$ are piecewise constant approximations of $H(y,T)$ and $G(y_i,t;y,T)$. In the algorithm, we only need the quantities $\{H(y_j,T)\}_{j\in\mathbb{Z}}$ and $\{G(y_i,t;y_j,T)\}_{j\in\mathbb{Z}}$. We use the full representations (10) and (11) in Section A of the supplementary online Appendix to prove Proposition 1. There, we show that relying on the sampling approximations (10) and (11) instead of the orthogonal projections given by the inner products as in Equation (9), leaves the convergence properties of the pricing algorithm unaffected.

B. *Merton-Black-Scholes: the recursive step*

The following proposition gives a recursion formula that relates the approximated values of the option at different points of the grid $\{y_j\}_{j=-\infty}^{\infty}$ and at different points in the set of exercise dates $\{t_l\}_{l=1,\ldots,L}$. It also states the rate of convergence of the algorithm.

Proposition 1: *Let $H(y,T)$ be such that $|H(y,T)-H(y',T)| < C|y-y'|$ for a positive constant C, and for $|y-y'| < \Delta y$. Furthermore, let $v_i(t_l)$ be defined for a set of dates $\{t_l\}_{l=1,\ldots,L}$, with $t_L = T$, as follows:*

$$v_i(t_l) = \max\Big\{H(y_i,t_l), \sum_{j\in\mathbb{Z}} G(y_i,t_l;y_j,t_{l+1})H(y_j,t_{l+1})\Delta y\Big\}, \quad \text{for} \quad l = L-1, \tag{12}$$

$$v_i(t_l) = \max\Big\{H(y_i,t_l), \sum_{j\in\mathbb{Z}} G(y_i,t_l;y_j,t_{l+1})v_j(t_{l+1})\Delta y\Big\}, \quad \text{for} \quad l = 1,\ldots,L-2. \tag{13}$$

Then, for each t_l in $\{t_1,\ldots,t_{L-1}\}$, the approximated values $v_i(t_l)$ defined in (12) *and* (13) *converge to the true value $V(y_i,t_l)$ with an approximation error of the order $O\big((\Delta y)^2\big)$.*

Proof. See Section A of the supplementary online Appendix. □

In principle, the continuity condition on $H(y,T)$ rules out digital payoffs. We need the condition to hold within each interval Δy of the grid $\{y_j\}_{j=-\infty}^{\infty}$. We can still price digital options by ensuring that the strike value belongs to the grid, and the convergence properties stated in Proposition 1 remain true. This procedure is not the same as placing

knots in the quadrature method because the grids remain the same for all dates $\{t_l\}_{l=1,\dots,L}$ provided that the strike price does not change with time. With a fixed grid, we can price even more exotic options, e.g. a digital call with a down-and-out feature, provided that the barrier occurs at the same value of the underlying for each t_l.

The main difference between Equations (12) and (13) is provided by the following explanation. In the right-hand side of (12), we find the exact values taken by the payoff function $H(y,T)$ on the grid $\{y_j\}_{j\in\mathbb{N}}$, and there is no approximation of the payoff. On the right-hand side of Equation (13), we find the values $\{v_j(t_{l+1})\}_{j\in\mathbb{N}}$ obtained in the previous step of the algorithm, and these are approximations of the true values $\{V(y_j,t_{l+1})\}_{j\in\mathbb{N}}$. Regardless of this fundamental difference, Proposition 1 states that the convergence rate is the same for both cases. In the case of a European option with maturity T, by taking $t=t_l$ in (12), we obtain the approximated price of the option at t. Equation (13) allows us to recursively compute the values of the option at different points in time, and thus to price Bermudan options, American options, and other types of path-dependent options.

In the implementation of Section II.A, we project the payoff function on a finite number of indicator functions. Thus, the values $\{H(y_j,T)\}_{j=1,\dots,N}$ and $\{\Delta y\, G(y_i,t;y_j,t_{l+1})\}_{j=1,\dots,N}$ are then the entries of the finite dimensions matrices $\mathbf{H}(T)$ and $\mathbf{G}(t_l;t_{l+1})$, respectively. Hence, we can express Equations (12) and (13) in exactly the same matrix forms as in (4) and (7). When the time step is a constant $\tau=t_{l+1}-t_l$, we obtain the fast and easily implementable algorithm of the introductory example. By choosing N sufficiently large, we can make the error introduced by truncating the infinite summations in (10) and (11) arbitrarily small. The truncation only suppresses the indicator functions localized in regions where the state price density vanishes, and where we can neglect the local contribution to the computed expectation.

Furthermore, as in Section II.B, whenever an exercise date t_l coincides with a dividend paying date t_h, we only need to replace the entries $\{\Delta y\, G(y_i,t_l;y_j,t_{l+1})\}_{j=1,\dots,N}$ of $\mathbf{G}(t_l;t_{l+1})$ with the values $\{\Delta y\, G(y_i-d,t_l;y_j,t_{l+1})\}_{j=1,\dots,N}$, whenever $t_l\in\{t_h\}_{h=1,\dots,H}$, to accommodate for a discrete dividend d.

C. Stochastic volatility: the projection step

In the class of stochastic volatility models, there are two state variables, the underlying asset S_t and the variance σ_t^2. The bivariate state price density $G_2(S_t,\sigma_t^2,t;y,w,T)$ describes the discounted transition probability density from the asset level S_t and variance level σ_t^2 at time t to the asset level $y=S_T$ and variance level $w=\sigma_T^2$ at time T. Its Fourier transform is denoted by $\hat{G}_2(S_t,\sigma_t^2,t;\lambda,\kappa,T)$, so that $G_2(S_t,\sigma_t^2,t;y,w,T)=\frac{1}{4\pi^2}\iint d\lambda d\kappa e^{-\iota(\lambda y+\kappa w)}\hat{G}_2(x,\xi,t;\lambda,\kappa,T)$, where ι is the imaginary unit.

For the values taken by the underlying asset, let the grid $\{y_j\}_{j\in\mathbb{N}}$ and the orthonormal set $\{e_j(y)\}_{j\in\mathbb{Z}}$ be defined as in the Black-Scholes case. The Fourier transforms of

$\{e_j(y)\}_{j\in\mathbb{Z}}$ are denoted by $\{\hat{e}_j(\lambda)\}_{j\in\mathbb{Z}}$, such that (see Section C of the supplementary online Appendix for the analytic form of $\hat{e}_j(y)$): $e_j(y) = \frac{1}{2\pi}\int d\lambda e^{-\iota\lambda y}\hat{e}_j(\lambda)$.

For the variance, we use the same equally spaced grid $\{w_q\}_{q\in\mathbb{N}}$ for the values taken by both the variables σ_t^2 and $w = \sigma_T^2$, such that $\Delta w = w_{q+1} - w_q$. Let $\{\varepsilon_q(w)\}_{q\in\mathbb{N}}$ be the normalized indicator functions centered on the grid $\{w_q\}_{q\in\mathbb{N}}$ and of width Δw, and $\{\hat{\varepsilon}_q(\kappa)\}_{q\in\mathbb{N}}$ be their Fourier transforms. Furthermore, let $\{\lambda_r\}_{r\in\mathbb{Z}}$ and $\{\kappa_z\}_{z\in\mathbb{Z}}$ be two regularly spaced grids of values taken by the transformed variables λ and κ, with constant widths $\Delta\lambda$ and $\Delta\kappa$, respectively.

The approximation of the payoff function $H(y,T)$ is the same as in Equation (10). Because we do not have the analytical form of $G_2(S_t,\sigma_t^2,t;y,w,T)$, we cannot directly sample the transition density, thus we must rely on the approximation given below. The projection step for the bivariate state price density $G_2(S_t,\sigma_t^2,t;y,w,T)$ when the conditioning variables take the values $S_t = y_i$ and $\sigma_t^2 = w_p$, is based on the following approximation:

$$\begin{aligned}
&\tilde{G}_2(y_i,w_p,t;y,w,T) \qquad (14)\\
&= \sqrt{\Delta y\Delta w}\sum_{j=-\infty,q=1}^{\infty}\Big(\frac{1}{4\pi^2}\sum_{r,z=-\infty}^{\infty}\hat{G}_2(y_i,w_p,t;\lambda_r,\kappa_z,T)\hat{e}_j(-\lambda_r)\hat{\varepsilon}_q(-\kappa_z)\Delta\lambda\Delta\kappa\Big)e_j(y)\varepsilon_q(w)\\
&\overset{\text{def}}{=} \sqrt{\Delta y\Delta w}\sum_{j=-\infty,q=1}^{\infty}\Gamma_2(y_i,w_p,t;y_j,w_q,T)e_j(y)\varepsilon_q(w),
\end{aligned}$$

where the second equality in (14) defines the quantities $\{\Gamma_2(y_i,w_p,t;y_j,w_q,T)\}_{j\in\mathbb{Z},q\in\mathbb{N}}$. To parallel the discussion following Equation (11) in the Black-Scholes case, the quantities $\{\Gamma_2(y_i,w_p,t;y_j,w_q,T)\}_{j\in\mathbb{Z},q\in\mathbb{N}}$ play the same role as $\{G(y_i,t;y_j,T)\}_{j\in\mathbb{Z}}$. We can motivate these approximations as follows. The orthogonal projection of the state price density $G_2(y_i,w_p,t;y,w,T)$ on the two orthonormal sets $\{e_j(y)\}_{j\in\mathbb{N}}$ and $\{\varepsilon_q(w)\}_{q\in\mathbb{N}}$ is given by the inner products $\iint dydw G_2(y_i,w_p,t;y,w,T)e_j(y)\varepsilon_q(w)$. Because we only know the closed form[4] of $\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)$ and not $G_2(y_i,w_p,t;y,w,T)$, we exploit the following key relationship:

$$\begin{aligned}
&\iint dydw G_2(y_i,w_p,t;y,w,T)e_j(y)\varepsilon_q(w)\\
&= \iint dydw\frac{1}{4\pi^2}\iint d\lambda d\kappa e^{-\iota(\lambda y+\kappa w)}\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)e_j(y)\varepsilon_q(w)\\
&= \frac{1}{4\pi^2}\iint d\lambda d\kappa\,\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)\int dy e^{-\iota\lambda y}e_j(y)\int dw e^{-\iota\kappa w}\varepsilon_q(w)\\
&= \frac{1}{4\pi^2}\iint d\lambda d\kappa\,\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)\hat{e}_j(-\lambda)\hat{\varepsilon}_q(-\kappa). \qquad (15)
\end{aligned}$$

Each $\Gamma_2(y_i,w_p,t;y_j,w_q,T)$ is an approximation of the last integral appearing in Equation (15), obtained by a direct sampling of the Fourier tranforms $\hat{G}_2(x_i,w_p,t;\lambda,\kappa,T)$, $\hat{e}_j(-\lambda)$

[4]For the closed form of $\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)$, see Griebsch (2013).

and $\hat{\varepsilon}_q(-\kappa)$ on the bivariate grid $\{(y_j, w_q)\}_{j\in\mathbb{Z},q\in\mathbb{N}}$, and on the univariate grids $\{\lambda_r\}_{r\in\mathbb{Z}}$ and $\{\kappa_z\}_{z\in\mathbb{Z}}$.

As in Section III.A, only the quantities $\{\Gamma_2(y_i, w_p, t; y_j, w_q, T)\}_{j\in\mathbb{N},q\in\mathbb{Z}}$ are the inputs for the pricing algorithm. The representation (14) is only used in Section B of the supplementary online Appendix to prove the convergence properties of the algorithm.

D. Stochastic volatility: the recursive step

In the stochastic volatility framework, the recursion for a Bermudan option consists of moving backwards in time as in Equation (6) with:

$$V(x,\xi,t_l) = \max\big\{H(x,t_l), \mathbb{E}\big[e^{-r(t_{l+1}-t_l)}V(S_{t_{l+1}},\sigma^2_{t_{l+1}},t_{l+1})\big|S_{t_l}=x,\sigma^2_{t_l}=\xi\big]\big\}. \tag{16}$$

Thus, the recursive step in the Heston model is the sampling counterpart of (16). The following proposition gives a recursion formula that relates the approximated values of the option at different points of the bivariate grid $\{(y_j, w_q)\}_{j\in Z,q\in\mathbb{N}}$ and at different points in time t_{l+1} and t_l. It also states the rate of convergence of the algorithm. Define $\Gamma_1(y_i, w_p, t_l; y_j, t_{l+1}) = \sum_{q=1}^{\infty}\Gamma_2(y_i, w_p, t_l; y_j, w_q, t_{l+1})\sqrt{\Delta w}$.

Proposition 2: *Let $H(y,T)$ be such that $|H(y,T) - H(y',T)| < C|y-y'|$ for a positive constant C, and for $|y-y'| < \Delta y$. Let $v_{ip}(t_l)$ be defined for a set of dates $\{t_l\}_{l=1,\ldots,L}$, with $t_L = T$, as follows:*

$$v_{ip}(t_l) = \max\Big\{H(y_i,t_l), \sum_{j=1}^{\infty}\Gamma_1(y_i,w_p,t_l;y_j,t_{l+1})H(y_j,t_{l+1})\sqrt{\Delta y}\Big\}, \qquad \text{for } l = L-1, \tag{17}$$

$$v_{ip}(t_l) = \max\Big\{H(y_i,t_l), \sum_{j,q=1}^{\infty}\Gamma_2(y_i,w_p,t_l;y_j,w_q,t_{l+1})v_{jq}(t_{l+1})\sqrt{\Delta y\Delta w}\Big\}, \text{ for } l = 1,\ldots,L-2. \tag{18}$$

Then, for each t_l in $\{t_1,\ldots,t_{L-1}\}$, the approximated values $v_{ip}(t_l)$ defined in (17) *and* (18) *converge to the true value $V(y_i,w_p,t_l)$ with an approximation error of the order $O(\underline{\Delta}^2)$, with $\underline{\Delta} = \sqrt{(\Delta y)^2 + (\Delta w)^2}$.*

Proof. See Section B of the supplementary online Appendix. □

For $t = t_{L-1}$, Equation (17) gives the price of a European option in the Heston model. Because the payoff function $H(y,T)$ only depends on the value y taken by the underlying asset at $t_L = T$, the computed price $v_{ip}(t_{L-1})$ depends on the stochastic variance only through the conditioning value $\sigma^2_{t_{L-1}} = w_p$. For this reason, we can use $\{\Gamma_1(y_i, w_p, t; y_j, T)\}_{j\in\mathbb{N}}$ instead of $\{\Gamma_2(y_i, w_p, t; y_j, w_q, T)\}_{j,q\in\mathbb{N}}$. Using the values $\{\Gamma_1(y_i, w_p, t; y_j, T)\}_{j\in\mathbb{N}}$ in (17) is equivalent to applying the projection step on the Fourier

transform $\hat{G}_1(y_i, w_p, t; \lambda, T) = \int d\kappa \hat{G}_2(y_i, w_p, t; \lambda, \kappa, T)$. It is the univariate function $\hat{G}_1$ and not the bivariate $\hat{G}_2$ that appears, for instance, in the original work by Heston (1993) for a European option. Figure 6 graphically presents the projection and recursion steps in the bivariate case.

[Figure 6 about here]

In the implementation, we truncate the summations in (17) and (18), so that the grid $\{(y_j, w_q)\}_{j=1,\ldots,N;q=1\ldots,W}$ has $N \times W$ points. The $N \times W$ matrix of computed prices at time $t = t_l$ is denoted by $\boldsymbol{v}_2(t_l)$, that is $\boldsymbol{v}_{2,jq}(t_l) = v_{jq}(t_l)$. Let $\boldsymbol{\Gamma}_2(y_i, w_p, t_l; t_{l+1})$ be the $N \times W$ matrix of the approximated transition probabilities from the initial point (y_i, w_p) to the end points of the entire grid $\{(y_j, w_q)\}_{j=1,\ldots,N;q=1\ldots,W}$. As in Section II.A, we integrate the normalization parameter $\sqrt{\Delta y \Delta w}$ in the definition of the transition matrix. We then have that $\boldsymbol{\Gamma}_{2,jq}(y_i, w_p, t_l; t_{l+1}) = \Gamma_2(y_i, w_p, t_l; y_j, w_q, t_{l+1})\sqrt{\Delta y \Delta w}$. Let $\phi_j = \{\hat{e}_j(-\lambda_r)\}_{r=1,\ldots,R}$ and $\varphi_q = \{\hat{\varepsilon}_q(-\kappa_z)\}_{z=1,\ldots,Z}$ be the values of the functions $\hat{e}_j(-\lambda)$ and $\hat{\varepsilon}_q(-\kappa)$ sampled at the grids $\{\lambda_r\}_{r=1,\ldots,R}$ and $\{\kappa_z\}_{z=1,\ldots,Z}$, respectively. Furthermore, we define the $R \times N$ matrix $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$, the $Z \times W$ matrix $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_W)$, and the $R \times Z$ matrix $\hat{\mathbf{G}}_2(y_i, w_p, t_l; t_{l+1})$ with entries $\hat{\mathbf{G}}_{2,rz}(y_i, w_p, t_l; t_{l+1}) = \hat{G}_2(y_i, w_p, t_l; \lambda_r, \kappa_z, t_{l+1})$. Then, we can write the coefficients of the projection step (14) in matrix form as:

$$\boldsymbol{\Gamma}_2(y_i, w_p, t_l; t_{l+1}) = \boldsymbol{\phi}' \hat{\mathbf{G}}_2(y_i, w_p, t_l; t_{l+1}) \boldsymbol{\varphi} \sqrt{\Delta y \Delta w}. \tag{19}$$

The recursive step (18) becomes the following:

$$\begin{aligned} v_{ip}(t_l) &= \max\Big\{ H(y_i, t_l), \sum_{j=1}^{N} \sum_{q=1}^{W} \Gamma_2(y_i, w_p, t_l; y_j, w_q, t_{l+1}) v_{jq}(t_{l+1}) \sqrt{\Delta y \Delta w} \Big\} \\ &= \max\big\{ H(y_i, t_l), \boldsymbol{\Gamma}_2(y_i, w_p, t_l; t_{l+1}) : \boldsymbol{v}_2(t_{l+1}) \big\}, \end{aligned} \tag{20}$$

where the symbol "**:**" denotes the Frobenius, or entry-wise, product.

In Section D of the supplementary online Appendix, we show how to speed up the computation of the matrices $\boldsymbol{\Gamma}_2(y_i, w_p, t_l; t_{l+1})$ by taking advantage of the space translation invariance property of transition densities. Our method contains the Fast Fourier Transform (FFT) as a special case. In the FFT, the univariate grids for λ and κ are automatically set, which can sometimes cause an imprecise reconstruction of the $\boldsymbol{\Gamma}_2(y_i, w_p, t_l; t_{l+1})$ matrices.

E. Numerical illustrations in the Heston model

We investigate the performance of our method in a standard affine model such as the Heston (1993) model. We study an American option, written on an asset S_t , which pays

discrete dividends and that evolves according to the following stochastic volatility model:

$$\begin{aligned} dX_t &= \Big(r - \frac{1}{2}\sigma_t^2\Big)dt + \sqrt{\sigma_t^2}\cdot dW_{1t}, \\ d\sigma_t^2 &= \beta\big(\sigma_{LT}^2 - \sigma_t^2\big)dt + \omega\sqrt{\sigma_t^2}\cdot dW_{2t}, \quad E(dW_{1t}\cdot dW_{2t}) = \rho dt. \end{aligned} \tag{21}$$

In Equation (21), $X_t = \log S_t$, and σ_t^2 is the variance process. We work with X_t to be able to implement the space invariance property of the transition matrices, as outlined in Section D of the supplementary online Appendix. We conduct two simulation studies. In the first, the American call has a time to maturity of one year, and 3 dividends worth $d = 2$ are distributed at $t_h = 0.25, 0.5, 0.75$. In the second, the time to maturity remains one year, but a single large dividend $d = 10$ is paid out after six months. The process parameter values are the following: $r = 0.05$, $\sigma_{LT} = 0.2$, $\beta = 2$ and $\omega = 0.2$. Moreover we choose the parameter ρ to be equal to zero. We compute the price for an at-the-money option ($S_0 = K = 100$). The benchmark method in this analysis is a finite-difference (hereafter *FD*) numerical solution of the partial derivatives equation (*PDE*) that describes the evolution of the price process V_t of the American call. We implement an alternating direction implicit (*ADI*) variant of the finite-difference scheme. For a recent discussion of schemes similar to *FD*, see, for instance, in't Hout and Foulon (2010). This implementation is equivalent to a Crank-Nicolson scheme, which in standard problems converges at a rate $O\big((\Delta t)^2\big)$, where Δt is the temporal discretization interval. In both the *FD* scheme and the recursive projections, the evolution of the option price V_t is studied on a rectangular grid in the space (X, σ^2), with $X \in [\log(K) - 10\sigma_{LT}\sqrt{T}, \log(K) + 10\sigma_{LT}\sqrt{T}]$ and $\sigma^2 \in [0, 0.3]$. In the *FD* scheme, the parameter m_s gives the number of equally spaced grid points in the X direction, and m_v gives the number of equally spaced grid points in the σ^2 direction, so that the grid points are $\{(X_i, \sigma_p^2)\}_{i=1,\ldots,m_s;p=1,\ldots,m_v}$. The parameter L_T gives the number of time steps used. In the recursive projections, under a sampling scheme we define $\Delta y = 2^{-J}a$, where a is a positive constant that gives the step of the $\{y_j\}_{j=1,\ldots,N}$ grid when $J = 0$. Describing the convergence of the recursive projections in terms of the parameter J emphasizes how the approximation error decreases each time the number of grid points is doubled. Similarly, $\Delta w = 2^{-J_w}a_w$, where a_w is the step with $J_w = 0$ of the $\{w\}_{p=1,..,W}$ grid in which the σ_t^2 variable takes values.

Assuming the contemporaneous correlation $\rho = 0$ simplifies the implementation of the *FD* scheme, in the sense that neglecting the correlation between X_t and σ_t^2 makes the *FD* scheme easier to code and faster. On the other hand, the speed and complexity of the recursive projection method are unaffected by the value chosen for the parameter ρ. The correlation is addressed in the Green function $G_2(x, \sigma_t^2, t; y, w, T)$ and consequently in the coefficients of the matrix $\hat{\mathbf{G}}_2$. Because the speed of the method depends on the number of entries in the $\hat{\mathbf{G}}_2$ matrix, and not on the values taken by the entries, it is clear that the choice of ρ does not affect the convergence rate of the recursive projections. This

feature is the first advantage of the recursive projection over finite-difference schemes. This simulation study will then give a lower bound to the difference in speed between the recursive projections and the *FD* scheme. To price an American option on dividend-paying stocks, we should implement the *FD* scheme-equivalent of the recombining tree. Doing so is practically unfeasible because it would mean computing at each ex-dividend date a new option price at each point of the grid. Instead, at each ex-dividend date t_h and at each grid point (X_i, σ_p^2), we opt for comparing the intrinsic value $H(X_i, t_h)$ with the continuation value $V(\widetilde{X}_i^d, \sigma_p^2, t_h)$, where $\widetilde{X}_i^d$ is the value of the X grid closest to $\log(e^{X_i} - d)$. This choice amounts to perturbating the *FD* scheme at each ex-dividend rate, which could translate into a convergence slower than the theoretical $O\big((\Delta t)^2\big)$. This feature is a second advantage of the recursive projection over the finite-difference schemes, because, as we explained in Section II.C, the recursive projections can easily adapt to discrete dividends without their affecting the convergence properties of the algorithm. The recursive projections achieve convergence quickly in the σ^2 direction. The method does not seem to improve by setting a resolution level greater than $J_w = 4$; thus, we keep this value fixed throughout our simulations. The *FD* scheme is also not very sensitive to the number of points used in the σ^2 direction. We find no improvement beyond $m_v = 31$.

Figure 7 shows the results for the 3-dividend case. The true value used to compute the pricing errors is 7.397, obtained with the resolution level $J = 13$. The graph on the right displays the pricing error of the *FD* scheme as a function of the time discretization parameter L_T. Each line is relative to a different value of the spatial discretization parameter m_s. The time labels are all relative to the $m_s = 3200$ curve. The *FD* scheme with $L_T = 2048$ and $m_s = 6400$ delivers a value within $1bp$; thus, we assume that the methods have converged when the absolute value of the relative error is within $1bp$ of 7.397. The graph on the left plots the relative pricing error of the recursive projections against the resolution level J. The regression line on the left graph shows that the estimated slope is almost exactly the slope of -2 predicted by the theoretical convergence results of Proposition 2. The *FD* is at least one order of magnitude slower. Compare, for instance, the computation time needed to deliver a $4bp$ error ($2s$ against $65s$), or a $1bp$ error ($8s$ against $130s$). Remember that we underestimate the speed advantages of the recursive projections in this experimental setup, with $\rho = 0$. Figure 8 compares the convergence speed of the two methods in the 1-dividend case. The true value of 7.302 is obtained by the recursive projection method with $J = 13$. The *FD* scheme requires 48 seconds to reach a $5bp$ relative error, with parameters $m_s = 400$ and $L_T = 2048$. The bottom curve, relative to $m_s = 200$, shows that the method does not converge for smaller values of the space discretization parameter. The small $5bp$ bias of the *FD* is due to the large value of the dividend d and the perturbation of the scheme at each dividend date. The rate as a function of the resolution level J at which the recursive projections attain the $1bp$ error band is approximately -2, as theoretically predicted.

Another notable difference between the *FD* and the recursive projection method is that the latter demands far fewer changes to adapt to different pricing problems. We have already observed, for example, in the Black-Scholes case that we can easily handle a digital payoff, both in the put case and in the call case. In Equation (4), the matrix $\mathbf{G}(t;T)$ depends only on the dynamics of the underlying asset and not on the payoff. We can compute it once for all and use it to price different options with different payoffs, because the payoff functional form only impacts the vector $\mathbf{H}(T)$. Such a design is particularly suited for object-oriented programming, which is often used in quant desks. In finite-difference schemes, the transition matrices are not independent of the payoff function, given that they depend on the conditions imposed on the function V_t on the boundaries of the rectangular grid on which it is computed. Options with different payoffs do not share the same boundary conditions and hence do not share the same transition matrices. It is feasible, as we actually did, to build a toolbox that receives the parameters of the dynamics and the parameters of the option contract and the payoff as inputs and that automatically generates the transition matrix and gives the prices as an output. A toolbox that, upon receiving the same inputs, automatically generates transition matrices in a finite-difference framework and uses them to price an option is simply unthinkable.

IV. Numerical applications and empirics

A. Numerical comparison of early exercise boundaries

In this section, we compare the early exercise boundary implied by the Black-Scholes model with those implied by the Heston stochastic volatility and the Merton jump-diffusion models. We study two cases in which i) the stock distributes a continuous dividend yield and ii) the stock distributes a discrete dividend. Combining cases i) and ii) with the different modelling assumptions for the underlying asset and different maturities leads to very different patterns. For instance, in the discrete dividend case, the early exercise boundary is lower under the Black-Scholes model than under the Heston model, whereas in the continuous dividend case, the opposite is true. Hence, by modelling a discrete dividend as a continuous yield, we can draw misleading conclusions in an empirical evaluation of suboptimal non-exercise. The exercise boundary S_t^* for an American call with a continuous dividend yield is defined as the lowest value of S_t such that $S_t - K \geq C(S_t, T, K)$. If the value of the current stock is above S_t^*, then it is optimal for the call holder to exercise his option. With a discrete dividend, it can only be optimal to exercise the call option on the days immediately before the ex-dividend dates t_h. The exercise boundary S_t^* for an American call option is then defined as the lowest value of S_{t_h} such that $S_{t_h} - K \geq C(S_{t_h} - d, T, K)$.

[Figure 9 about here]

In Panel A of Figure 9, we plot the early exercise boundary for the Heston and Black-Scholes models for an American call option with a continuous dividend yield $r_d = 0.03$ (right graph) and with an equivalent quarterly discrete dividend $d = 1.38$ (left graph). We choose $d = 1.38$ to have an equivalent total annual dividend between the continuous dividend yield $r_d = 0.03$ and the discrete dividend case. Indeed, $1.38 = 0.03S^*/4$, where $S^* = 184$ is the critical stock price under the Black-Scholes model in the dividend yield case for maturity $T = 0.5$. We use the following set of representative parameters: $T = 0.5$, $K = 100$, $r = 0.05$, $\sigma_0 = 0.2$, $\omega = 0.1$, $\sigma_{LT} = 0.3$, $\beta = 4$, and $\rho = -0.5$ (Adolfsson et al. (2013)). For comparison, we follow Heston (1993), and we use the Black-Scholes model with a volatility parameter that matches the (square root of the) variance of the spot return over the life of the option in the Heston model. When the stock distributes a regular quarterly dividend, there are only two dividend payments during the life time of an option with maturity $T = 0.5$, and it is immediately before the payment dates that it can be optimal to exercise the option. In our example, the two dividend payments occur immediately, at $t = 0$, and at $t = 0.25$, corresponding to a time to maturity of 0.5 and 0.25, respectively. At both dates, the value of the exercise boundary is lower under the Black-Scholes model. With a continuous dividend yield, the Heston early exercise boundary is always below the Black-Scholes boundary, whereas with discrete dividends, the opposite is the case.

Although the findings in the continuous dividend case are in line with those of Adolfsson et al. (2013), the findings in the discrete dividend case are entirely new. This difference warrants further intuitive discussion. Assume there is only one discrete dividend to be paid. The continuation value of the call option immediately after the ex-dividend date is that of a European call with the remaining time to maturity. When $\rho \leq 0$, the price of European options for a deep in-the-money call, where early exercise could be optimal, is higher in the Heston case than in the Black-Scholes case (see Heston (1993); Hull and White (1987)). For instance, in the left graph of Panel B of Figure 9, for a time to maturity of 0.25, this would be the case in the range of stock prices of approximately 150. Even by taking into account the dividend drop in computing the continuation value, the ex-dividend stock price should remain in the region where the Heston price is higher. We can repeat the same argument for a number of discrete dividends sufficiently small (typically of the order of a couple percent) to prevent the stock price from falling in the price range where the call has more value under the Black-Scholes model. The behavior of the boundary with a continuous dividend is less straightforward to grasp. Following Kim (1990) and Jamshidian (1992), we can decompose the value $V(S_t, t)$ of an American option into two components, namely, the European value $V^E(S_t, t)$ and the early exercise

premium $V^A(S_t, t)$, such that:

$$\begin{aligned} V(S_t, t) &= V^E(S_t, t) + V^A(S_t, t) \\ &= e^{-r(T-t)}\mathbb{E}\big[(S_T - K)_+ \big| S_t, \sigma_t^2\big] + \int_t^T e^{-r(s-t)}\mathbb{E}\big[(r_d S_s - rK)\mathbb{I}_{(S_s > S_s^*)} \big| S_t, \sigma_t^2\big] ds, \end{aligned} \tag{22}$$

where S_s^* is the early exercise boundary at time s and $\mathbb{I}_{(S_s > S_s^*)}$ equals one if, at time s, the stock is in the exercise region, otherwise zero. We can interpret $V^A(S_t, t)$ as a continuum of European call options with maturity $T - s$, strike price S_s^*, and payoff $r_d S_s - rK$. For each of these European options, we can apply the results of Table II in Hull and White (1987) who compare the values of European options under general stochastic volatility dynamics with the Black-Scholes price. Call values under the stochastic volatility assumption are lower when the contracts are at-the-money and $\rho \leq 0$. The continuum of contracts composing the $V^A(S_t, t)$ are at-the-money when $S_s = S_s^*$. As confirmed from our numerical simulations, the S_s^* values are distributed in the region immediately above $S = 150$, that is, exactly where the price of the American option under the Heston model is lower than that under the Black-Scholes model, and explain the negative bump in the right graph of Panel B of Figure 9.

Similarly, we can characterize the early exercise under the Merton jump-diffusion model, where the asset S_t evolves according to the following jump-diffusion process:

$$\frac{dS_t}{S_t} = (r - r_d - \gamma\upsilon)dt + \sigma_M dW_t + (\psi - 1)dq_t, \tag{23}$$

where r_d is the continuous dividend yield paid by the asset, and σ_M^2 is the instantaneous variance of the return conditional on no jump arrivals. The Poisson process, $q(t)$, is independent of W_t, and such that there is a probability γdt that a jump occurs in dt, and $1 - \gamma$ probability that no jump occurs. The parameter γ represents the mean number of jumps per unit of time. The random variable ψ is such that $\psi - 1$ describes the percentage change in the stock price if the Poisson event occurs, and $\upsilon = \mathbb{E}[\psi - 1]$ is the mean jump size. We further make the standard assumption (for instance, see Amin (1993); Bakshi et al. (1997)) that $\log(\psi) \sim N(\mu_\psi, \sigma_\psi^2)$. If $\gamma = 0$, then we recover the standard Black-Scholes model with no jumps. We use the following set of representative parameters: $K = 40$, $r = 0.08$, $\gamma = 5$, $\sigma_M^2 = 0.05$, $\sigma_\psi^2 = 0.05$, $\mu_\psi = 0$ (Amin (1993)). We set the volatility parameter in the Black-Scholes model equal to the volatility of the underlying return over the life of the option in the Merton model.

[Figure 10 about here]

In the Panel A of Figure 10, we plot the early exercise boundary for the Merton and Black-Scholes models for an American call option with a continuous dividend yield $r_d = 0.05$ (right graph) and in the case in which the stock pays an equivalent quarterly

discrete dividend $d = 1.125$ [5] (left graph). As for the Heston case, the results on the continuous dividend case are in line with the existing literature, e. g., Amin (1993), and we provide new insights into the discrete dividend case. To interpret the graphs in Figure 10, we have to make an important distinction. For short maturity options, the jump component in Equation (23) dominates the diffusion component. As explained in Amin (1993) and Merton (1976), the result is higher prices for short maturities under the Merton model than under the Black-Scholes model. We call this effect the jump effect. For longer maturities, the jump effect no longer dominates the diffusion component but instead creates an interplay that makes the jump-diffusion process observationally similar to a stochastic volatility process. For a discrete dividend, both the jump effect and the stochastic volatility effect, as previously discussed in the Heston case, predict a higher boundary in the Merton case than in the Black-Scholes case, which holds true for all maturities. This result is exactly what we find in the left graph of Panel A of Figure 10. For a continuous dividend case, we have that the jump effect dominates for short maturities, giving a higher boundary under the Merton model than under the Black-Scholes model. For longer maturities, the jump effect diminishes, and the boundary behaves as in the stochastic volatility model, that is, taking lower values than in the Black-Scholes case. These insights explain the crossing of the early exercise boundary that we observe in the right graph of Panel A, Figure 10.

A key numerical finding of this section is that the early exercise is more likely under the Black-Scholes model when we face discrete dividends. In the next section, we assess the empirical consequences of this finding for the cost of suboptimal non-exercise.

B. Empirical analysis

In this section, we apply the recursive projection method to characterize the early exercise boundary of a large sample of call options. The options have a maturity of less than six months and are written on dividend-paying stocks, which are part of the Dow Jones Industrial Average Index (DJIA). The sample comprises daily observations between January 1996 and December 2012. We investigate the early exercise decision of call holders in light of the different values that the exercise boundary can take under distinct modelling assumptions for the underlying asset. Following the procedure suggested by Pool et al. (2008), we first check which contracts should be exercised by comparing the intrinsic value immediately before the dividend payment with the continuation value on the ex-dividend day. We quantify how much is economically lost in the case of a suboptimal non-exercise decision. This amount depends on the continuation value and is model-specific. We compare the results obtained under three modelling environ-

[5] As before, we take $d = 1.125$ because $1.125 = 0.05S^*/4$, where $S^* = 90$ is the critical stock price with the dividend yield $r_d = 0.05$ for maturity $T = 0.5$.

ments, namely, Black-Scholes, Merton jump-diffusion, and Merton jump-diffusion with the stochastic volatility dynamics of the Heston model. Bates (1996) was the first to suggest combining the Merton and Heston models, and therefore, we refer to this process specification as the Bates model. Finally, whenever we find evidence of a suboptimal non-exercise decision, we show that trading costs alone cannot justify the behavior of investors.

In our empirical analysis, we price by fully taking into account the discrete nature of the dividend distributed by the underlying stocks and the American style of the call options, and we do so for the three pricing models. This feature is a peculiarity of our work, given that the standard empirical literature on options mainly focuses on European $S\&P500$ options with a dividend yield or limits itself to the Black-Scholes model. Our choice for alternative modelling environments follows the empirical findings of Bakshi et al. (1997), who suggest that jumps and stochastic volatility play a dominant role in pricing short-term options whereas modelling stochastic interest rates does not seem to significantly improve the pricing performance. In addition, the choice of the jump arrival distribution in Equation (23) is motivated by the work of Bajgrowicz, Scaillet, and Treccani (2015), who show that high-frequency data on individual stocks support the hypothesis that jump arrivals follow a simple low-intensity Poisson process. Their findings also support the assumption in Merton (1976) that the jump component is nonsystematic, i.e., diversifiable, because of the absence of co-jumps affecting all stocks. This explains our choice of the Bates model, where $X_t = \log(S_t)$, with:

$$\begin{aligned} dX_t &= (r - r_d - \gamma\upsilon - \frac{1}{2}\sigma_t^2)dt + \sigma_t dW_{1,t} + log(\psi)dq_t, \\ d\sigma_t^2 &= \beta\big(\sigma_{LT}^2 - \sigma_t^2\big)dt + \omega\sqrt{\sigma_t^2}\cdot dW_{2,t}, \quad E(dW_{1,t}\cdot dW_{2,t}) = \rho dt, \end{aligned} \tag{24}$$

and $\log(\psi) \sim N(\mu_\psi, \sigma_\psi^2)$. To implement the recursive projections in the Bates model, we need the Green function implied by Equation (24). Because the jump process is independent of the Brownian motions $dW_{1,t}$ and $dW_{2,t}$, we use the property that the characteristic function of a bivariate process is the product of the characteristic functions of the independent univariate processes. The transition matrices are then obtained as in Section III.D. The Bates model shows how straightforward it is to adapt the recursive projections to more complex models. The number of entries in the transition matrices for the Bates model is exactly the same as it is as for the Heston model. In the calibrations, we have compared the computation of the transition matrices with both the FFT and the full sampling method outlined in Section III.C. Because we have found that the FFT correctly reproduces the $\mathbf{\Gamma}_2(y_i, w_p, t_l; t_{l+1})$ matrices, we have opted for using the FFT throughout the empirical exercise to take advantage of the further increase in speed provided by the FFT algorithm.

The daily data on all option attributes, the stock price, and the dividend distribution

details are from Optionmetrics. We obtain the daily data on the interest rates from the Treasury constant maturities of the H15 report of the Federal Reserve[6]. A total of 101,295 series of short-term options written on 30 stocks enter our database. The total number of records is approximately 9.5 million. This number stresses the importance of a fast and versatile numerical method. Table I reports the number of quotes for each stock with a breakdown by maturity and moneyness. Our study focuses on the early exercise behavior of investors; hence, we focus on the in-the-money options, which are the category of options for which the number of quotes is the highest.

B.1. Estimating the cost of suboptimal non-exercise

Table II reports the results of the calibration for the three modelling frameworks. We obtain the parameters through the minimization of the implied volatility mean squared error, as in Christoffersen and Jacobs (2004)[7]. The first line of Table II displays the average values of the parameters calibrated on our sample of single stocks, whereas the second line reports the average values that Bakshi et al. (1997) obtain for the same parameters from contracts written on the $S\&P500$ index. The parameters that rule the level of the volatility smile, namely, the Black-Scholes volatility σ_{BS}, the long term volatility σ_{LT}, and the spot volatility σ_0, are much higher in our single stock calibration than in the index calibration, which reflects the well known fact that an index is less volatile than its components. Indeed, in our sample, the average Black-Scholes volatility is 29%, σ_0 is 28%, and the average long-term implied volatility is 32%, whereas for the index options, the same parameters take the values of 18.15%, 20%, and again 20%. The jump parameters in the Bates model show that jumps are on average less frequent in the single stock case than in the index case ($\gamma_{Stocks} = 0.5$ against $\gamma_{SP500} = 0.61$), but the amplitude and variability are higher for single stocks ($\mu_{\psi,Stocks} = -0.12$ and $\sigma_{\psi,Stocks} = 0.18$ respectively) than for the index ($\mu_{\psi,SP500} = -0.09$ and $\sigma_{\psi,SP500} = 0.14$). Given that the index is a diversified portfolio, it displays a jump whenever one of its constituents jumps, but the non-systematic nature of the jumps of single stocks (see Bajgrowicz et al. (2015)) attenuates the variability and magnitude at an aggregate level. The remaining two parameters of the stochastic volatility component of the Bates model, the correlation parameter ρ and the volatility of volatility ω, have a specific impact on the shape of the implied volatility smile (Hagan, Kumar, Lesniewski, and Woodward (2002); West (2005)). A negative ρ implies a negatively sloped smile. The correlation parameter is in absolute value lower in the single stock case ($\rho_{SP500} = -0.52$ against $\rho_{Stocks} = -0.35$), meaning that the implied volatility smile for the index is more negatively sloped than

[6]We access the Optionmetrics and the H15 databases through the Wharton Research Data Services (WRDS) research platform.

[7]Section F of the supplementary online Appendix gives a detailed description of the data and the calibration procedure.

for individual stocks. This finding is consistent with the findings of Bakshi, Kapadia, and Madan (2003) who describe the same relationship between the slopes of the index and of the individual stocks implied volatility smiles. Bollen and Whaley (2004) also find the same pattern, and explain it by relating the slope of the index smile to the buying pressure for index puts, with the demand for call options driving the shape of the smile of single stocks. The volatility of volatility ω determines the convexity of the implied volatility smile. The difference in the values taken by ω is striking. We find ω to be 75% for stock options, whereas Bakshi et al. (1997) find a much smaller value of 40% for short-term index options. This difference is due to the higher convexity of the implied volatility smiles of stock options versus that of index options, another feature also documented in Bollen and Whaley (2004). The parameters ρ and ω are related to the smile shape through the higher moments of the distributions of the returns of the underlying. A more negative ρ generates a more negatively skewed distribution of index returns with respect to stock returns, whereas a higher ω in the single stock returns leads to a higher kurtosis than in the index return distribution.

In Table I of the supplementary online Appendix, we provide more detailed results, including a breakdown of the calibration by stock, and we show that the values of the calibrated parameters are homogeneous across stocks.

[Table II about here]

After having calibrated the models, we are able to compute the price $C(S_{t_h-1}-d, K, T)$ on the day previous to the dividend payment date t_h, by using the recursive projections. The price $C(S_{t_h-1} - d, K, T)$ is the continuation value of the option at date t_h, when the dividend d is distributed. By comparing it with the intrinsic value $S_{t_h-1} - K$, we can assess which options should be exercised on $t_h - 1$. If an option should be exercised (i.e., $C(S_{t_h-1} - d, K, T) \leq S_{t_h-1} - K$), then a positive open interest at the end of the day before ex-dividend ($OI_{t_h-1} > 0$) measures the failure of investors to exercise the option. In this case, we calculate the suboptimal non-exercise percentage as the following ratio:

$$NE_{\%} = \frac{OI_{t_h-1}}{OI_{t_h-2}}, \tag{25}$$

i.e., the number of contracts outstanding at the end of the day $t_h - 1$ to the total number of contracts outstanding at the end of day $t_h - 2$. The quantity defined in Equation (25) is an approximation of the actual non-exercise ratio, because it neglects a possible issue of new contracts on date $t_h - 1$. This event is unlikely; indeed Pool et al. (2008) test the approximation on a subsample of contracts for which they have the real exercise data. They conclude that the approximation (25) is very precise. The total amount of money that is left on the table due to suboptimal non-exercise is given by the following formula:

$$TL = 100 \times OI_{t-1} \times [(S_{t-1} - K) - C(S_{t-1} - d, K, T)]. \tag{26}$$

The continuation value $C(S_{t-1} - d, K, T)$ depends on the model used for pricing; hence, the total loss due to suboptimal non-exercise (TL) is itself model-specific.

Table IV presents the results on the suboptimal non-exercise behavior of investors.

[Table IV about here]

Table IV clearly shows that the optimal early exercise decision depends on the model used for the stock price. Under the Black-Scholes model, approximately 9.5% of the outstanding contracts should be exercised, and the percentage decreases (approximately 7.5%) under the alternative models. This result is consistent with the numerical findings of Section IV.A, where we show that, in the case of discrete dividends, the early exercise boundary under the Black-Scholes model is lower compared to that implied by the Merton and Heston models. As a general rule, contracts that should be exercised under the Merton or Bates models should also be exercised under the Black-Scholes model. In our sample, we find some exceptions to this rule because, in Section IV.A, we choose the model parameters such that the total variance of the returns over the life of the option is the same in all models, whereas in real data, this condition may not hold. To give some examples, 4680 options should be exercised under Black-Scholes but not under the Bates model, whereas the reverse is true only for 249 contracts. Similarly, we find that 2872 options should be exercised under the Black-Scholes model but not under the Merton model, whereas the opposite occurs with only 53 options. The first important lesson we learn is that, by allowing for more sophisticated models than the Black-Scholes model, the number of contracts that should be optimally exercised decreases by almost 25%. The suboptimality figures are model-dependent and may be a consequence of the calibration procedure. The comparison between our calibration results and the results of Bakshi et al. (1997) are reassuring in terms of the reliability of our calibration method. To justify the suboptimal behavior found in our sample, we should obtain unreasonably high values for the jump and intensity parameters.

A second piece of evidence that stands out from Table IV is that the percentage of investors who leave the options suboptimally non-exercised is higher under the Black-Scholes model than under the other models, 39% versus approximately 30%. We compute these percentages in accordance with Definition (25). If we restrict our attention to the 1965 contracts in our sample that should be exercised under the Black-Scholes model but not under the Merton model or the Bates model, we find a striking 81% of no-exercise. These results may suggest that investors do not limit themselves to a Black-Scholes world when evaluating their options but rely on more sophisticated models that include jumps or stochastic volatility. Even if this evidence is a considerable step towards understanding the investor decision-making process, it does not fully solve the puzzle. Indeed, even in the Merton and Bates models, we still find a high percentage of suboptimal non-exercises, which leads to a global loss of approximately $130-140$ million dollars, down approximately 30% from the loss of 206 million dollars in the Black-Scholes model.

A second possible explanation of the early exercise puzzle is that investors wait until the options are deeper in-the-money before exercising. If we restrict ourselves to the options that should be exercised only under the Black-Scholes model but not under the Merton or Bates models, then the average moneyness of the contracts, measured by the delta of the options, is 0.962. If we consider all the contracts that should be exercised under the Black-Scholes model, including those that should also be exercised under the other two model specifications, the average moneyness rises to 0.986. Figure 11 shows how relevant the moneyness of the contract is in the non-exercise decision of investors: the more in-the-money the option is, the smaller the number of suboptimal non-exercises. Investors may not respond immediately to favorable stock price movements and may take some time before reacting and optimally exercising their option, which would be in line with the behavior on the early prepayment of mortgages documented by Stanton (1995).

Our summary results regarding the exercise decision in the Black-Scholes model are in line with those obtained by Pool et al. (2008). In that work, the authors apply the early exercise decision rule to all options series by using the Black-Scholes model with historical volatility and find that 53.1% of investors leave their options unexercised when instead they should have been exercised. Their data span over ten years (from 1996 to 2006) and to compare our results with theirs, we divide our sample into two subsamples, the first spanning the years 1996-2006 and the second spanning the years 2006-2012. Then, we calculate the average percentage of suboptimal non-exercise in the two subsamples and find that the percentage of suboptimal non-exercise under the Black-Scholes model is approximately 47% in the first subsample and 37% in the second. The decrease in the non-exercise behavior with time intimates that investors become more attentive in monitoring their investments. There is a small difference between our results (47%) and the 53.1% found in Pool et al. (2008). The explanation is most likely our focus on the constituents of the Dow Jones Industrial Average, whereas Pool et al. (2008) consider all option series. It is likely that, for large-cap companies, stock and option prices are monitored more closely than they are on average.

Throughout our empirical investigation, we choose a model-based approach to calculate the continuation value of the option $C(S_{t_h-1} - d, K, T)$. We could have also used a market-based approach where the continuation value is the market price of the option. The market-based approach checks whether the quantity $C_{MKT}(S_{t_h-1}, K, T) - (S_{t_h-1} - K)_+$ equals 0, where $C_{MKT}(S_{t_h-1}, K, T)$ is the observed market price at $t = t_h - 1$. As discussed in Pool et al. (2008) and in Barraclough and Whaley (2012), the market-based approach has shortcomings. The most important is that it does not make it possible to calculate the total loss due to suboptimal non-exercise, which we do in Equation (26). In addition, the bid-ask spread and the discreteness of the prices make it difficult to decide which C_{MKT} should be used. For all of these reasons, we follow Pool et al. (2008) and Barraclough and Whaley (2012) and use a model-based approach. Barraclough and

Whaley (2012) only use the market-based approach as a useful model-free test to confirm the presence of suboptimal non-exercise behavior. They find that the market-based approach gives a magnitude of suboptimality that is comparable to that implied by the model-based approach. This last piece of evidence is an additional argument against the possible objection that an incorrect calibration of the model parameters may be the source of the suboptimal exercise figures.

B.2. The role of fees

According to the recent literature on option prices (Jensen and Pedersen (2014); Christoffersen, Goyenko, Jacobs, and Karoui (2014)), trading costs and financial frictions in general strongly affect both the option prices and the early exercise decision of American options. In this section, we investigate whether the suboptimal non-exercise behavior of investors is due to the trading costs that investors face when exercising their options.

Following Pool et al. (2008), we model the costs of exiting a long call position as a per share lump sum $\mathscr{F}$ that the investor must pay at the moment he decides to exercise. The specific value of $\mathscr{F}$ depends on how the exit is accomplished according to the different possible objectives of the investor. The most expensive value of the fee $\mathscr{F}$ is attained when the investor wants to exercise the option and reenter into the same call position. Pool et al. (2008) estimate an average value for the rollover costs $\mathscr{F}$ by using the commissions of the high-cost brokers, and they obtain a very conservative amount of $\mathscr{F} = 0.4446$ dollar per share. A detailed description of the components of the fee $\mathscr{F}$ can be found in Pool et al. (2008).

To understand the role of the fees in the early exercise decision, we perform two different empirical exercises. As a first check, we re-perform the exercise of Section IV.B.1 and compute the loss due to a suboptimal non-exercise decision, but this time using $C(S_{t_h-1} - d, K + \mathscr{F}, T)$ as the continuation value, and $(S_{t_h-1} - K - \mathscr{F})$ as the intrinsic value. The fee value is $\mathscr{F} = 0.4446$. The fee $\mathscr{F}$ enters both in the exercise proceeds and in the continuation value. Indeed, at the moment of the exercise decision, the investor should decide whether to exercise and hence pay the exercise fee immediately or not exercise and postpone the payment of the exercise fee to a future date. Accordingly, the calculation of the total amount of money that is lost due to suboptimal non-exercise is given by the following formula:

$$TL_{\mathscr{F}} = 100 \times OI_{t-1} \times [(S_{t-1} - K - \mathscr{F}) - C(S_{t-1} - d, K + \mathscr{F}, T)]. \tag{27}$$

The second column in Table IV shows the summary results including the fee. They are not very different from those obtained without considering the fee (first column of Table IV). We can conclude that the inclusion of trading costs does not change the big picture

on the suboptimal non-exercise of investors, as outlined in the previous paragraph.

As a second empirical exercise, we calculate the value of the fee that would justify the non-exercise decision of investors to detect possible additional costs that are not taken into consideration in the fee $\mathscr{F}$. To do so, for each option for which $C(S_{t_h-1} - d, K, T) < (S_{t_h-1} - K)$, but that is not optimally exercised by some of the investors, we compute the value of the implied fee $\mathscr{F}$ that would justify the non-exercise decision. It amounts to numerically finding the zero of the following function:

$$f(\mathscr{F}) = C(S_{t_h-1} - d, K + \mathscr{F}, T) - (S_{t_h-1} - K - \mathscr{F}). \tag{28}$$

The results are reported in Table III. The average implied fee is between 7 and 8 dollars per share, an incredibly high amount compared to the already conservative fee of 0.4446 dollar per share estimated by Pool et al. (2008). No realistic hidden fees can sum up to 7 dollars per share, and the trading costs of exiting a long call option position cannot fully justify the suboptimal non-exercise behavior of investors. We can interpret the difference between the implied fee of 8 dollars and the conservative fee of 0.4446 as an implied opportunity cost for the holder of the option to monitor the optimal exercise of the American option. The holder of the option chooses to spend this amount on alternative activities.

V. Concluding remarks

In this paper, we introduce a fast and accurate method for pricing a large class of path-dependent options, those for which the path is being monitored at discrete moments of time. We start from the observation that, by monitoring the value of an option at discrete times and by sampling the value function of the contract on a finite grid of values of the underlying asset, we can describe the evolution of the price process in terms of elementary matrix operators. The interpretation of the elements of such a matrix in terms of a functional projection allows us to extend the matrix approach to the pricing of contracts written on assets following processes whose transition probabilities do not have a known analytical expression in the direct space. The stochastic volatility process and variance gamma processes are examples of such frameworks. The recursive projection method owes its speed to the simplicity of its algorithm, which is based on sampling. The speed and accuracy of the method allows us to study problems that are typically difficult to address due to numerical complexity. Our recursive projection allows pricing American options on assets that receive dividends at discrete points in time.

As a notable application, we characterize the early exercise boundary of an American call option on a dividend-paying stock under different modelling assumptions. We find that the boundary is higher under the Merton and Heston models than under the Black-Scholes model if the dividend is discrete and that it is lower in the case of a con-

tinuous dividend yield. These differences underline the importance of the model and dividend distribution assumptions in option pricing. In the empirical study, we calibrate the Black-Scholes, Merton and Bates models on a sample of American call option prices and examine the early exercise behavior of long American call option holders. In accordance with our numerical results, we find that the cost of failing to optimally exercise is higher under the Black-Scholes model than it is under the Merton and Bates models. We can envisage two lines of research for further investigation: other possible applications of our method and improving the algorithm. American option pricing is only a particular case of stochastic optimal control problems. We can think of applying the recursive projection method to other problems, such as the optimal portfolio allocation involving complex and path-dependent financial assets. Currently, these types of complex problems are solved by using Monte Carlo simulations (Detemple, Garcia, and Rindisbacher (2003)) and our method could offer a more efficient computational alternative. The second possible extension of our work is the use of different function bases in the projection step. Candidate basis systems are functions that share the localization property with the indicator function but that may display higher regularity and enhance the convergence speed. Faster convergence should not come at the cost of a more complicated numerical implementation, and therefore, research should go in the direction of functional projections that result in linear transformations of the sampled values (Sweldens (1996, 1998)).

REFERENCES

Adolfsson, Thomas, Carl Chiarella, Andrew Ziogas, and Jonathan Ziveyi, 2013, Representation and numerical approximation of American option prices under Heston stochastic volatility dynamics, Research Paper 327, Quantitative finance research centre, University of Technology Sidney, available at http://www.qfrc.uts.edu.au/research/research_papers/rp327.pdf.

Aït-Sahalia, Yacine, and Robert L. Kimmel, 2007, Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics* 83, 413–452.

Aït-Sahalia, Yacine, and Robert L. Kimmel, 2010, Estimating affine multifactor term structure models using closed-form likelihood expansions, *Journal of Financial Economics* 98, 113–144.

Amin, Kaushik I., 1993, Jump diffusion option valuation in discrete time, *The Journal of Finance* 48, 1833–1863.

Andersen, Leif, and Mark Broadie, 2004, Primal-dual simulation algorithm for pricing multidimensional American options, *Management Science* 50, 1222–1234.

Andricopoulos, Ari D., Martin Widdicks, Peter W. Duck, and David P. Newton, 2003, Universal option valuation using quadrature methods, *Journal of Financial Economics* 67, 447 – 471.

Andricopoulos, Ari D., Martin Widdicks, David P. Newton, and Peter W. Duck, 2007, Extending quadrature methods to value multi-asset and complex path dependent options, *Journal of Financial Economics* 83, 471–499.

Bajgrowicz, Pierre, Olivier Scaillet, and Adrien Treccani, 2015, Jumps in high-frequency data: Spurious detections, dynamics, and news, *Management Science* .

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *The Journal of Finance* 52, 2003–2049.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Barone-Adesi, Giovanni, and Robert E. Whaley, 1987, Efficient analytic approximation of American option values, *The Journal of Finance* 42, 301–320.

Barraclough, Kathryn, and Robert E. Whaley, 2012, Early exercise of put options on stocks, *The Journal of Finance* 67, 1423–1456.

Bates, David S., 1996, Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, *Review of Financial Studies* 9, 69–107.

Battauz, Anna, Marzia De Donno, and Alessandro Sbuelz, 2014, Real options and American derivatives: The double continuation region, *Management Science* 61, 1094–1107.

Ben-Ameur, Hatem, Michèle Breton, and Juan-Manuel Martinez, 2009, Dynamic programming approach for valuing options in the GARCH model, *Management Science* 55, 252–266.

Bollen, Nicolas P. B., and Robert E. Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *The Journal of Finance* 59, 711–753.

Brennan, Michael J., and Eduardo S. Schwartz, 1977, The valuation of American put options, *The Journal of Finance* 32, pp. 449–462.

Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2009, Understanding index option returns, *Review of Financial Studies* 22, 4493–4529.

Broadie, Mark, and Jerome Detemple, 1996, American option valuation: New bounds, approximations, and a comparison of existing methods, *Review of Financial Studies* 9, 1211–1250.

Broadie, Mark, and Paul Glasserman, 1997, Pricing American-style securities using simulation, *Journal of Economic Dynamics and Control* 21, 1323–1352.

Bunch, David S., and Herb Johnson, 2000, The American put option and its critical stock price, *The Journal of Finance* 55, 2333–2356.

Carr, Peter, Hélyette Geman, Dilip B. Madan, and Marc Yor, 2003, Stochastic volatility for Lévy processes, *Mathematical Finance* 13, 345–382.

Carr, Peter, Robert Jarrow, and Ravi Myneni, 1992, Alternative characterizations of american put options, *Mathematical Finance* 2, 87–106.

Chen, Ding, Hannu J. Härkönen, and David P. Newton, 2014, Advancing the universality of quadrature methods to any underlying process for option pricing, *Journal of Financial Economics* 114, 600–612.

Cheng, Peng, and Olivier Scaillet, 2007, Linear-quadratic jump-diffusion modeling, *Mathematical Finance* 17, 575–598.

Chiarella, Carl, Nadima El-Hassan, and Adam Kucera, 1999, Evaluation of american option prices in a path integral framework using Fourier–Hermite series expansions, *Journal of Economic Dynamics and Control* 23, 1387–1424.

Chiarella, Carl, and Andrew Ziogas, 2005, Pricing american options on jump-diffusion processes using Fourier Hermite series expansions, *Quantitative Finance Research Centre Research Paper* .

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2014, Illiquidity premia in the equity options market, *Available at SSRN 1784868* .

Christoffersen, Peter, and Kris Jacobs, 2004, The importance of the loss function in option valuation, *Journal of Financial Economics* 72, 291–318.

Clarke, Nigel, and Kevin Parrott, 1999, Multigrid for American option pricing with stochastic volatility, *Applied Mathematical Finance* 6, 177–195.

Cox, John C., Stephen A. Ross, and Mark Rubinstein, 1979, Option pricing: A simplified approach, *Journal of Financial Economics* 7, 229–263.

Darolles, Serge, and Jean-Paul Laurent, 2000, Approximating payoffs and pricing formulas, *Journal of Economic Dynamics and Control* 24, 1721–1746.

Desai, Vijay V., Vivek F. Farias, and Ciamac C. Moallemi, 2012, Pathwise optimization for optimal stopping problems, *Management Science* 58, 2292–2308.

Detemple, Jérôme, 2005, *American-style derivatives: Valuation and computation*, Financial mathematics series (Chapman & Hall/CRC Press, Boca Raton, Florida, USA).

Detemple, Jerome, and Weidong Tian, 2002, The valuation of american options for a class of diffusion processes, *Management Science* 48, 917–937.

Detemple, Jerome B, Rene Garcia, and Marcel Rindisbacher, 2003, A Monte Carlo method for optimal portfolios, *The Journal of Finance* 58, 401–446.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.

Eraker, Bjorn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *The Journal of Finance* 58, 1269–1300.

Garman, Mark B., 1985, Towards a semigroup pricing theory, *The Journal of Finance* 40, 847–861.

Geske, Robert, 1979, A note on an analytical valuation formula for unprotected American call options on stocks with known dividends, *Journal of Financial Economics* 7, 375–380.

Geske, Robert, and H. E. Johnson, 1984, The american put option valued analytically, *The Journal of Finance* 39, 1511–1524.

Griebsch, Susanne A, 2013, The evaluation of European compound option prices under stochastic volatility using Fourier transform techniques, *Review of Derivatives Research* 16, 135–165.

Hagan, Patrick S., Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward, 2002, Managing smile risk, *WILMOTT Magazine* 84–108.

Hansen, Lars Peter, and José A. Scheinkman, 2009, Long-term risk: An operator approach, *Econometrica* 77, 177–234.

Haugh, Martin B., and Leonid Kogan, 2004, Pricing American options: a duality approach, *Operations Research* 52, 258–270.

Heston, Steven L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–343.

Huang, Jing-zhi, Marti G. Subrahmanyam, and G. George Yu, 1996, Pricing and hedging American options: a recursive investigation method, *Review of Financial Studies* 9, 277–300.

Hull, John, and Alan White, 1987, The pricing of options on assets with stochastic volatilities, *The Journal of Finance* 42, 281–300.

Ibáñez, Alfredo, 2003, Robust pricing of the american put option: A note on Richardson extrapolation and the early exercise premium, *Management Science* 49, 1210–1228.

Ikonen, Samuli, and Jari Toivanen, 2008, Efficient numerical methods for pricing American options under stochastic volatility, *Numerical Methods for Partial Differential Equations* 24, 104–126.

in't Hout, Karel J., and S. Foulon, 2010, ADI finite difference schemes for option pricing in the Heston model with correlation, *International Journal of Numerical Analysis and Modeling* 7, 303–320.

Jackson, Kenneth R., Sebastian Jaimungal, and Vladimir Surkov, 2008, Fourier space time-stepping for option pricing with Lévy models, *Journal of Computational Finance* 12, 1–29.

Jamshidian, Farshid, 1992, An analysis of American options, *Review of Futures Markets* 11, 72–80.

Jensen, Mads Vestergaard, and Lasse Heje Pedersen, 2014, Early option exercise: Never say never, Working paper, Copenhagen Business School, Copenhagen.

Ju, Nengjiu, 1998, Pricing an American option by approximating its early exercise boundary as a multipiece exponential function, *Review of Financial Studies* 11, 627–646.

Kim, IN, 1990, The analytic valuation of American options, *Review of Financial Studies* 3, 547–572.

Kristensen, Dennis, and Antonio Mele, 2011, Adding and subtracting Black-Scholes: a new approach to approximating derivative prices in continuous-time models, *Journal of Financial Ecomonics* 102, 390–415.

Lacoste, Vincent, 1996, Wiener chaos: a new approach to option hedging, *Mathematical Finance* 6, 197–213.

Lamberton, Damien, and Stéphane Villeneuve, 2003, Critical price near maturity for an american option on a dividend-paying stock, *The Annals of Applied Probability* 13, pp. 800–815.

Leippold, Markus, and Liuren Wu, 2002, Asset pricing under the quadratic class, *The Journal of Financial and Quantitative Analysis* 37, 271–295.

Linetsky, Vadim, 1997, The path integral approach to financial modeling and options pricing, *Computational Economics* 11, 129–163.

Longstaff, Francis A., and Eduardo S. Schwartz, 2001, Valuing American options by simulation: a simple least-squares approach, *Review of Financial Studies* 14, 113–147.

Lord, Roger, Fang Fang, Frank Bervoets, and Cornelis W. Oosterlee, 2008, A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes, *SIAM Journal on Scientific Computing* 30, 1678–1705.

Madan, Dilip B., Peter P. Carr, and Eric C. Chang, 1998, The variance gamma process and option pricing, *European Finance Review* 2, 79–105.

Madan, Dilip B, and Frank Milne, 1994, Contingent claims valued and hedged by pricing and investing in a basis, *Mathematical Finance* 4, 223–245.

Medvedev, Alexey, and Olivier Scaillet, 2010, Pricing American options under stochastic volatility and stochastic interest rates, *Journal of Financial Economics* 98, 145–159.

Merton, Robert C., 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Ecomonics* 3, 125–144.

Miller, Merton H., and Kevin Rock, 1985, Dividend policy under asymmetric information, *The Journal of Finance* 40, 1031–1051.

Ødegaard, Bernt Arne, 2014, Financial numerical recipes in C++., Unpublished manuscript, University of Stavanger.

O'Sullivan, Conall, 2005, Path dependant option pricing under Lévy processes, in *EFA 2005 Moscow Meetings Paper. Available at SSRN: http://ssrn.com/abstract=673424* .

Pool, Veronika Krepely, Hans R. Stoll, and Robert E. Whaley, 2008, Failure to exercise call options: An anomaly and a trading game, *Journal of Financial Markets* 11, 1–35.

Rogers, Leonard C. G., 2002, Monte Carlo valuation of American options, *Mathematical Finance* 12, 271–286.

Roll, Richard, 1977, An analytic valuation formula for unprotected American call options on stocks with known dividends, *Journal of Financial Economics* 5, 251–258.

Stanton, Richard, 1995, Rational prepayment and the valuation of mortgage-backed securities, *Review of Financial Studies* 8, 677–708.

Sullivan, Michael A., 2000, Valuing American put options using gaussian quadrature, *Review of Financial Studies* 13, 75–94.

Sweldens, Wim, 1996, The lifting scheme: A custom-design construction of biorthogonal wavelets, *Applied and Computational Harmonic Analysis* 3, 186–200.

Sweldens, Wim, 1998, The lifting scheme: A construction of second generation wavelets, *SIAM Journal on Mathematical Analysis* 29, 511–546.

Vellekoop, Michel H., and J. W. Nieuwenhuis, 2006, Efficient pricing of derivatives on assets with discrete dividends, *Applied Mathematical Finance* 13, 265–284.

West, Graeme, 2005, Calibration of the SABR model in illiquid markets, *Applied Mathematical Finance* 12, 371–385.

Whaley, Robert E., 1981, On the valuation of American call options on stocks with known dividends, *Journal of Financial Economics* 9, 207–211.

Xiu, Dacheng, 2014, Hermite polynomial based expansion of European option prices, *Journal of Econometrics* 179, 158–177.

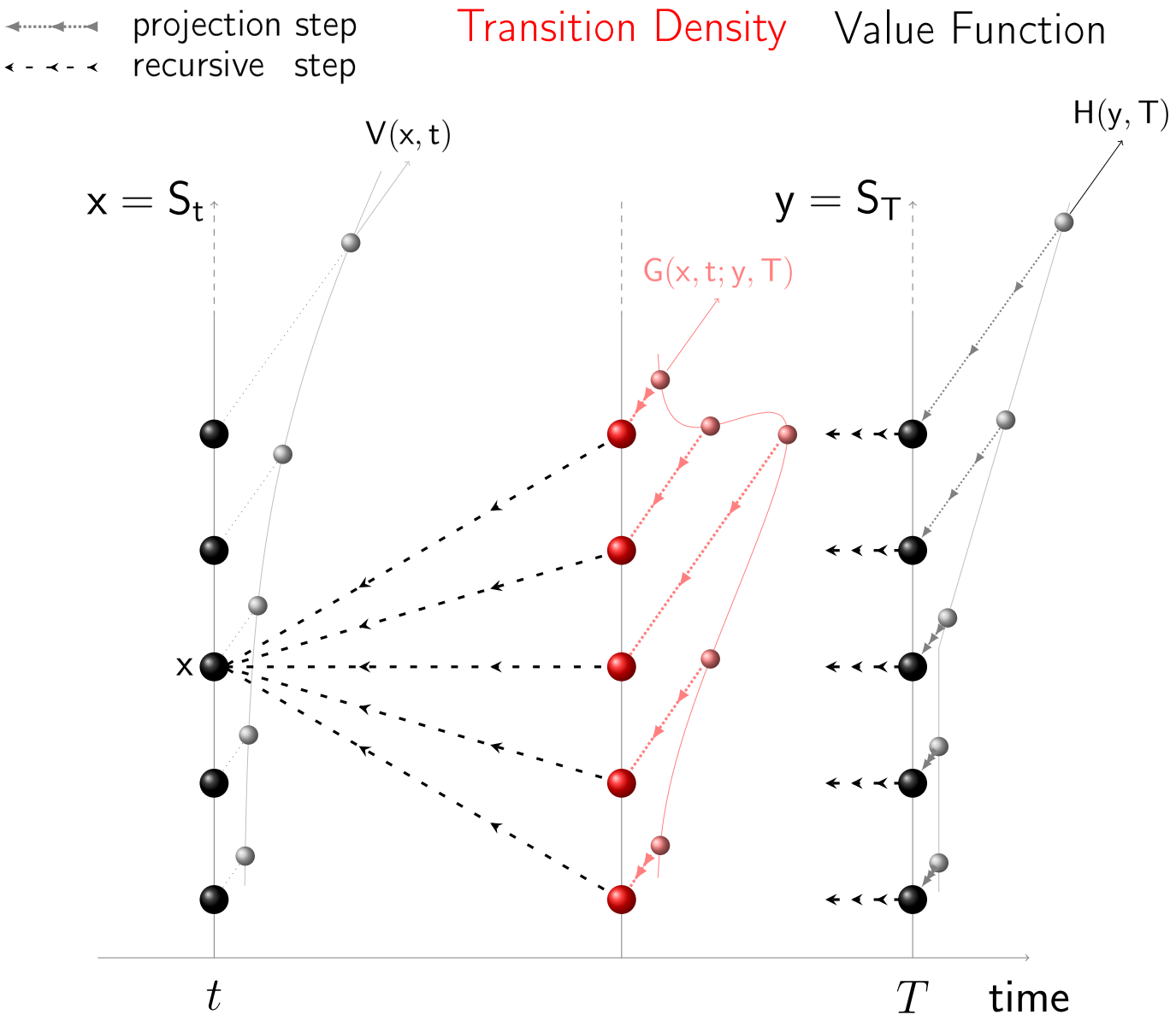

Figure 1: Recursive Projection scheme in the Black-Scholes case is composed of two steps. First the projection step: the value function $H(y,T)$ at T and state price density functions $G(x,t;y,T)$ are sampled (*dotted lines*). Second the recursive step: the sampled values are multiplied by the transition weights to obtain the value function at t (*dashed lines*), which in turn will be the input for the following step of the algorithm.

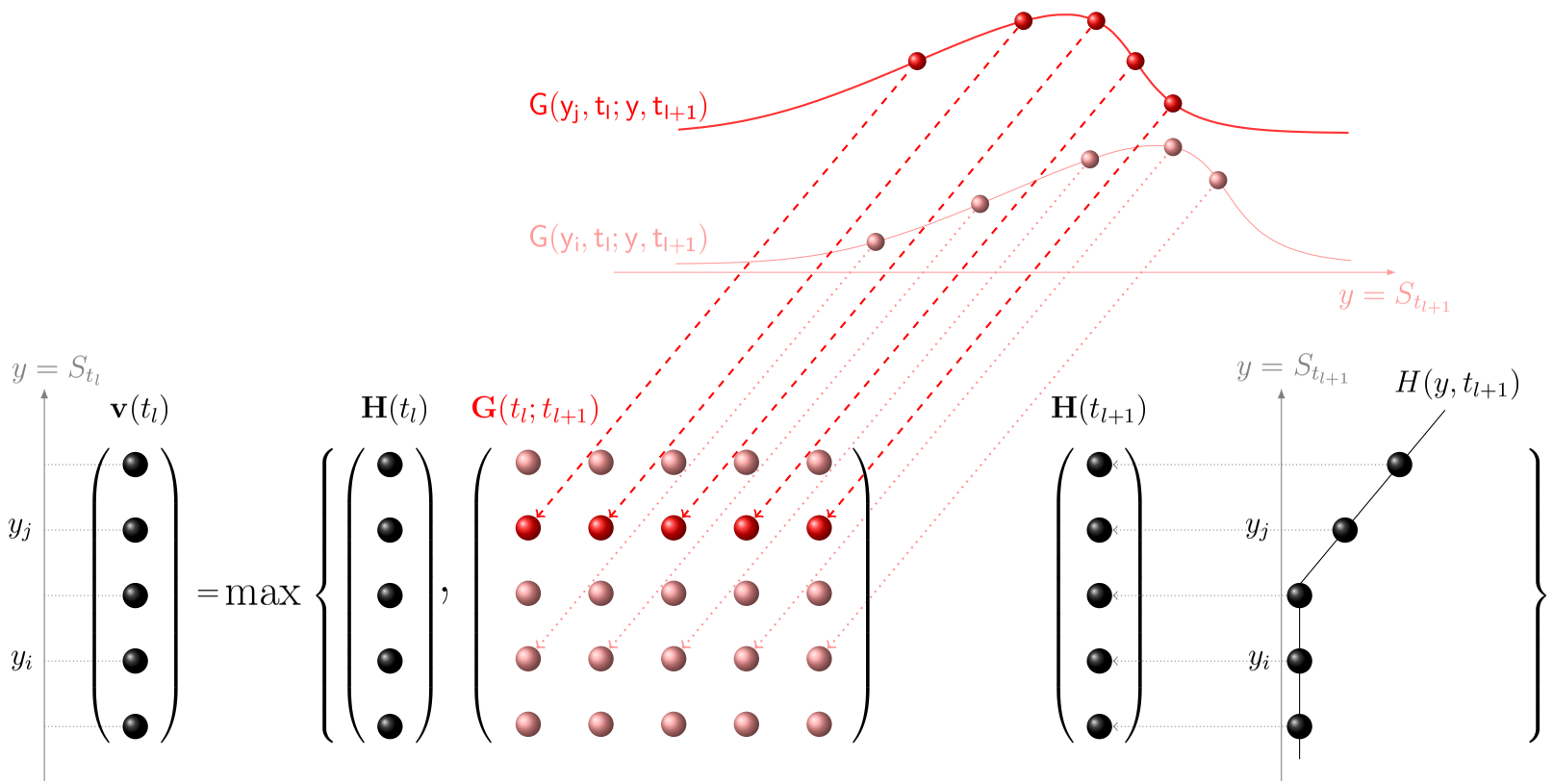

Figure 2: Recursive Projection scheme in the Black-Scholes case when grids are the same at each exercise date t_l. The rows of the $\mathbf{G}(t_l; t_{l+1})$ matrix are given by the values sampled from the transition densities $G(y_i, t_l; y_j, t_{l+1})$ for conditioning values $\{y_1, \ldots, y_N\}$. In this way, the output of the vector times matrix multiplication contains the continuation values at the grid points $\{y_1, \ldots, y_N\}$. Each continuation value at t_l is compared with the intrinsic value $H(y_i, t_l)$, to obtain the vector $\mathbf{v}(t_l)$, whose elements are the approximation of $V(y, t_l)$ at the same grid points $\{y_1, \ldots, y_N\}$, and that is the input for the following recursive step.

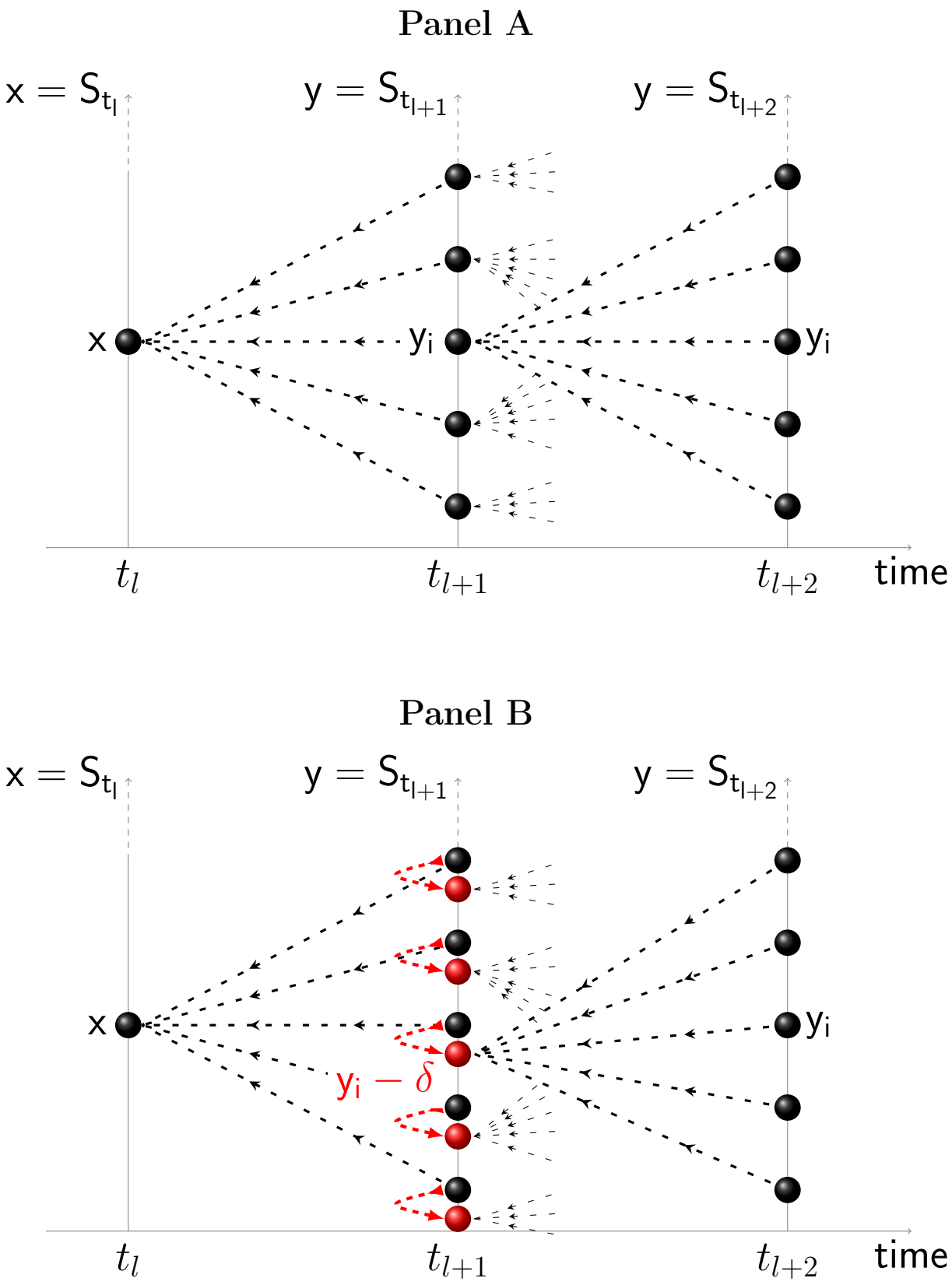

Figure 3: Recursive scheme without dividends (Panel A) and with discrete dividends (Panel B). In Panel A, at date $t = t_{l+1}$, the intrinsic value $H(y_i, t_{l+1})$ is compared with the continuation value $V(y_i, t_{l+1})$ computed at the same grid point y_i (black ball). In Panel B, at the ex-dividend date $t_h = t_{l+1}$, the intrinsic value $H(y_i, t_{l+1})$ at the grid point y_i (black ball) is compared with the continuation value $V(y_i - d, t_{l+1})$ at $y_i - d$ (red ball).

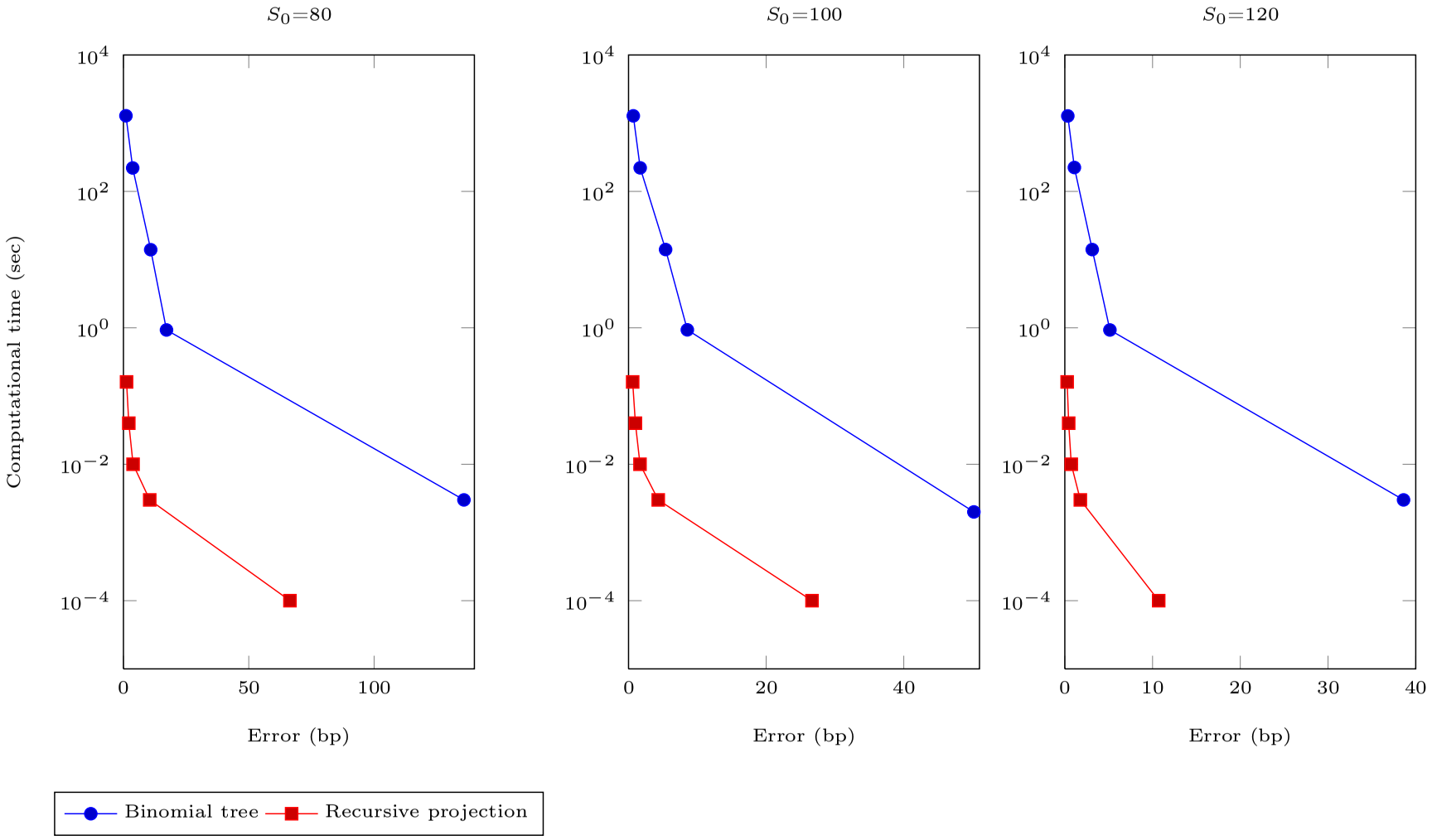

Figure 4: Comparison between the binomial tree and the recursive projection method on an American call option written on a dividend paying stock in the Black-Scholes case. The option has a maturity of 3 years and a dividend $d = 2$ is paid at the end of each year. Other parameters are set equal to $r = 0.05$, $\sigma = 0.2$, $K = 100$.

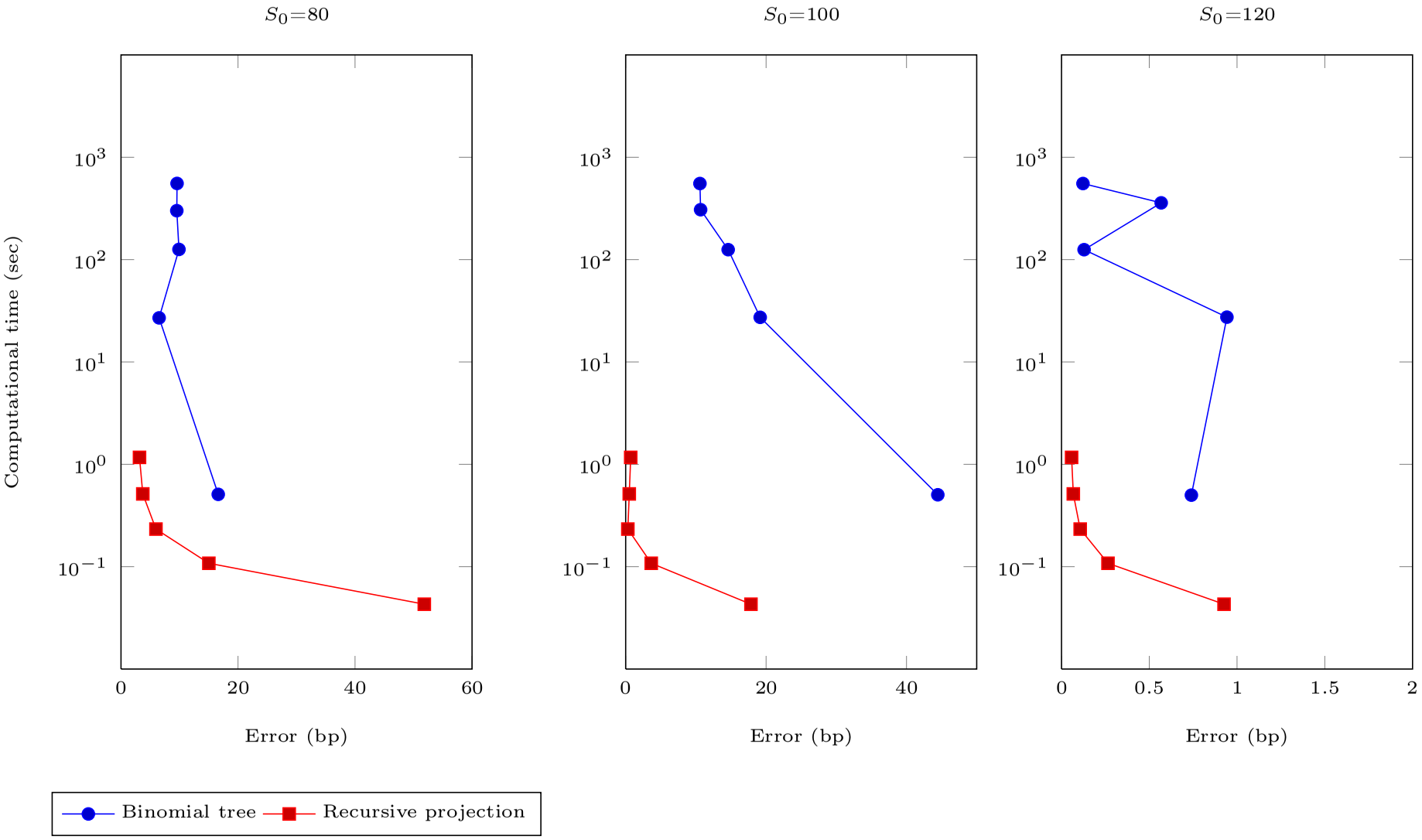

Figure 5: Comparison between the binomial tree and the recursive projection method on a Bermudan digital call option in the Black-Scholes case. The option has a maturity of 10 years and can be exercised 4 times per year. Other parameters are set equal to $r = 0.1$, $\sigma = 0.2$, and $K = 100$.

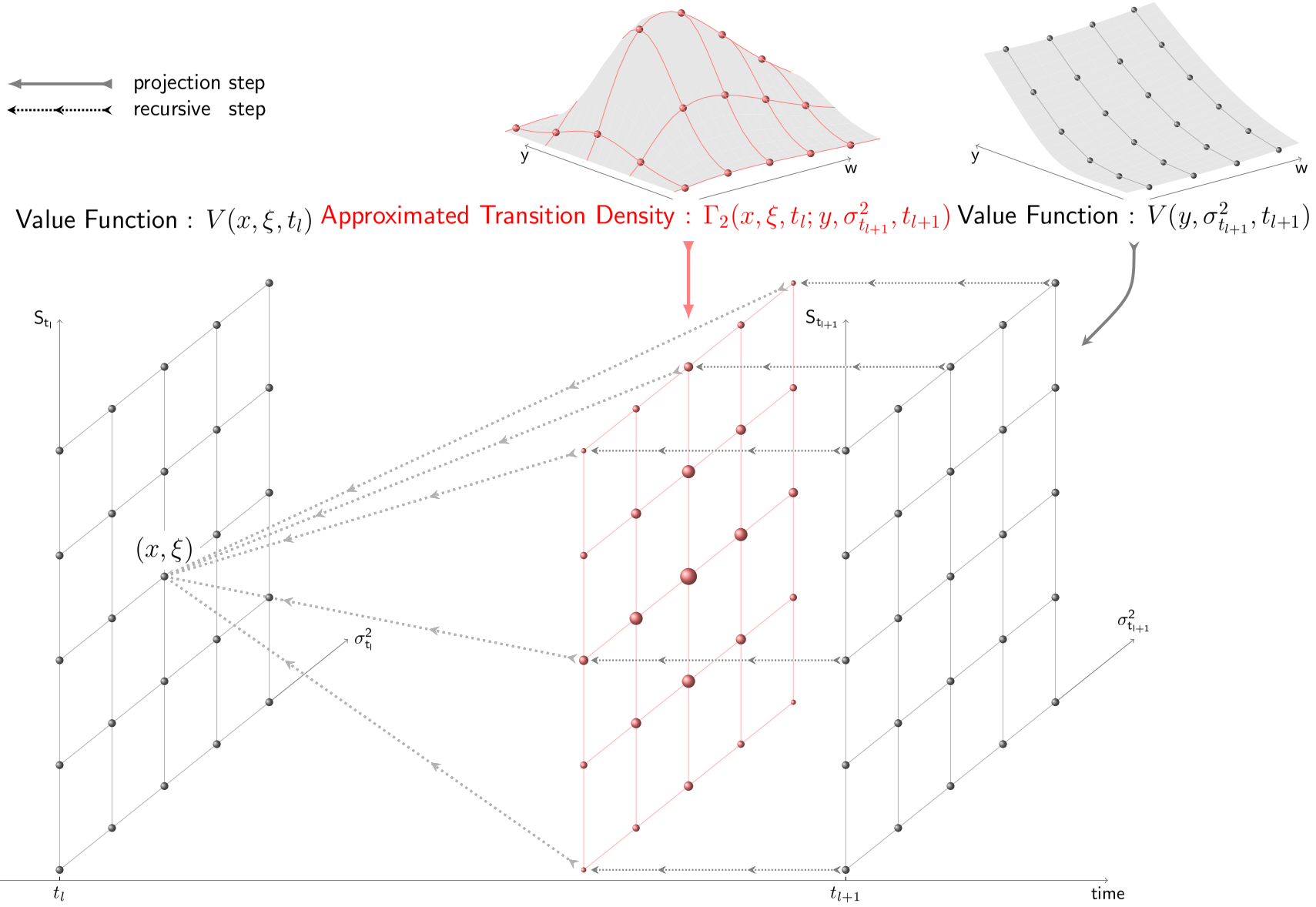

Figure 6: Recursive Projection scheme in the stochastic volatility case is composed of two steps. First the projection step (*thick arrows*): the value function $V(y, w, t)$ at $t = t_{l+1}$ is sampled and the state price density function $G_2(x, \xi, t_l; y, w, t_{l+1})$ is approximated at (y_j, w_q) by $\Gamma_2(x, \xi, t_l; y_j, w_q, t_{l+1})$. Second the recursive step: the sampled values are multiplied by the transition weights to obtain the value function $V(x = y_i, \xi = w_p, t_l)$ at $t = t_l$ (*thin arrows*), which in turn will be the input for the following step of the algorithm.

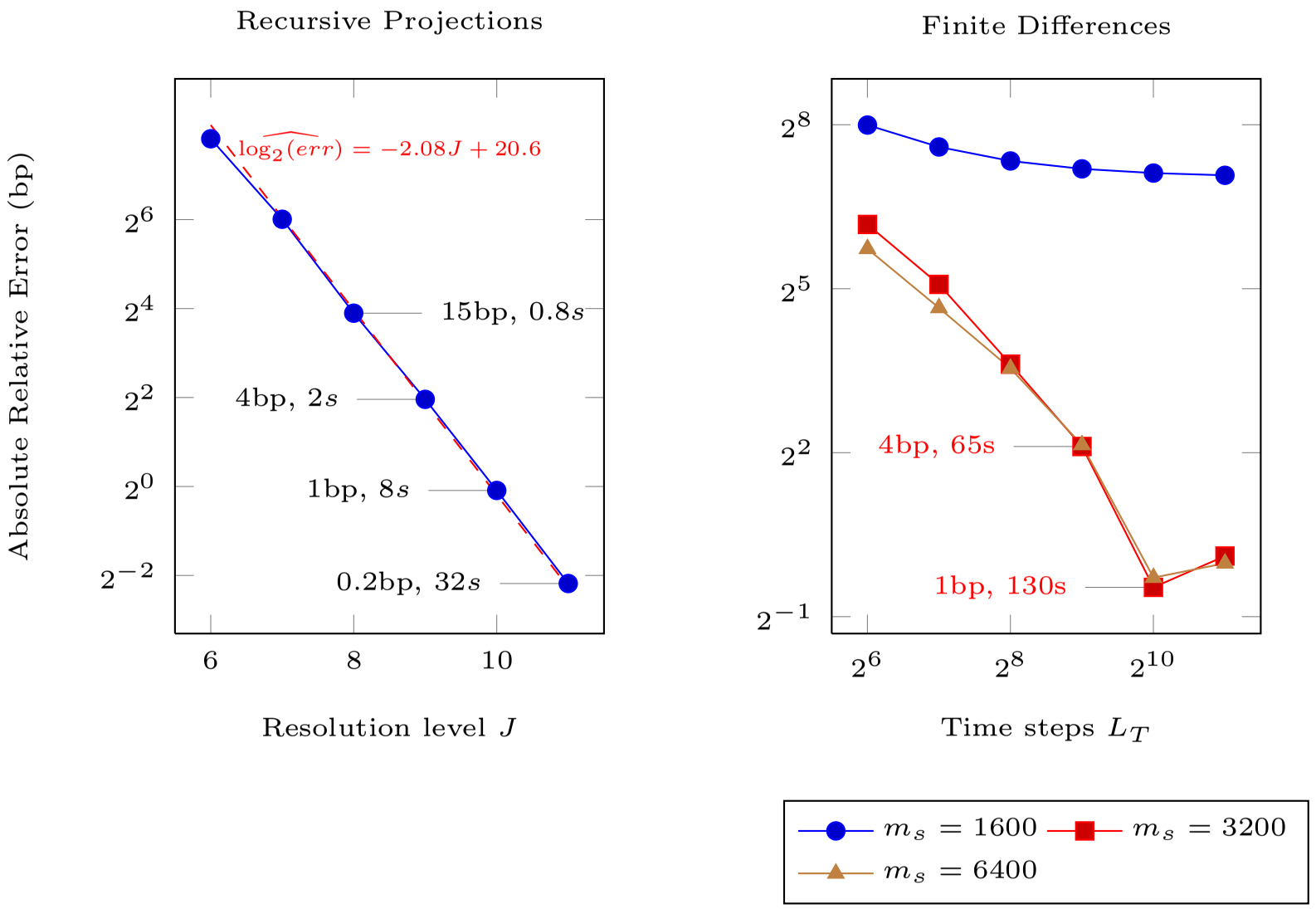

Figure 7: Comparison between the finite difference scheme and the recursive projections on an American call option written on a dividend-paying stock in the Heston case. The option has a maturity of 1 year and a dividend $d = 2$ is paid at dates $t_l = 0.25, 0.5, 0.75$. Other parameters are set equal to $S_0 = 100$, $K = 100$, $r = 0.05$, $\sigma_{LT} = 0.2$, $\beta = 2$, $\omega = 0.2$. The parameter m_s gives the umber of points in the $X_t = \log S_t$ grid for the *FD* scheme, while the σ^2 grid has $m_w = 31$ points. The sampling grid for the $X_t = \log S_t$ variable in the recursive projections has size 2^J. The resolution level in the σ^2 dimension is $J_w = 4$. The dashed line in the left panel corresponds to a fitted linear regression, and shows that the estimated slope is close to the slope of -2 predicted by the theoretical convergence results.

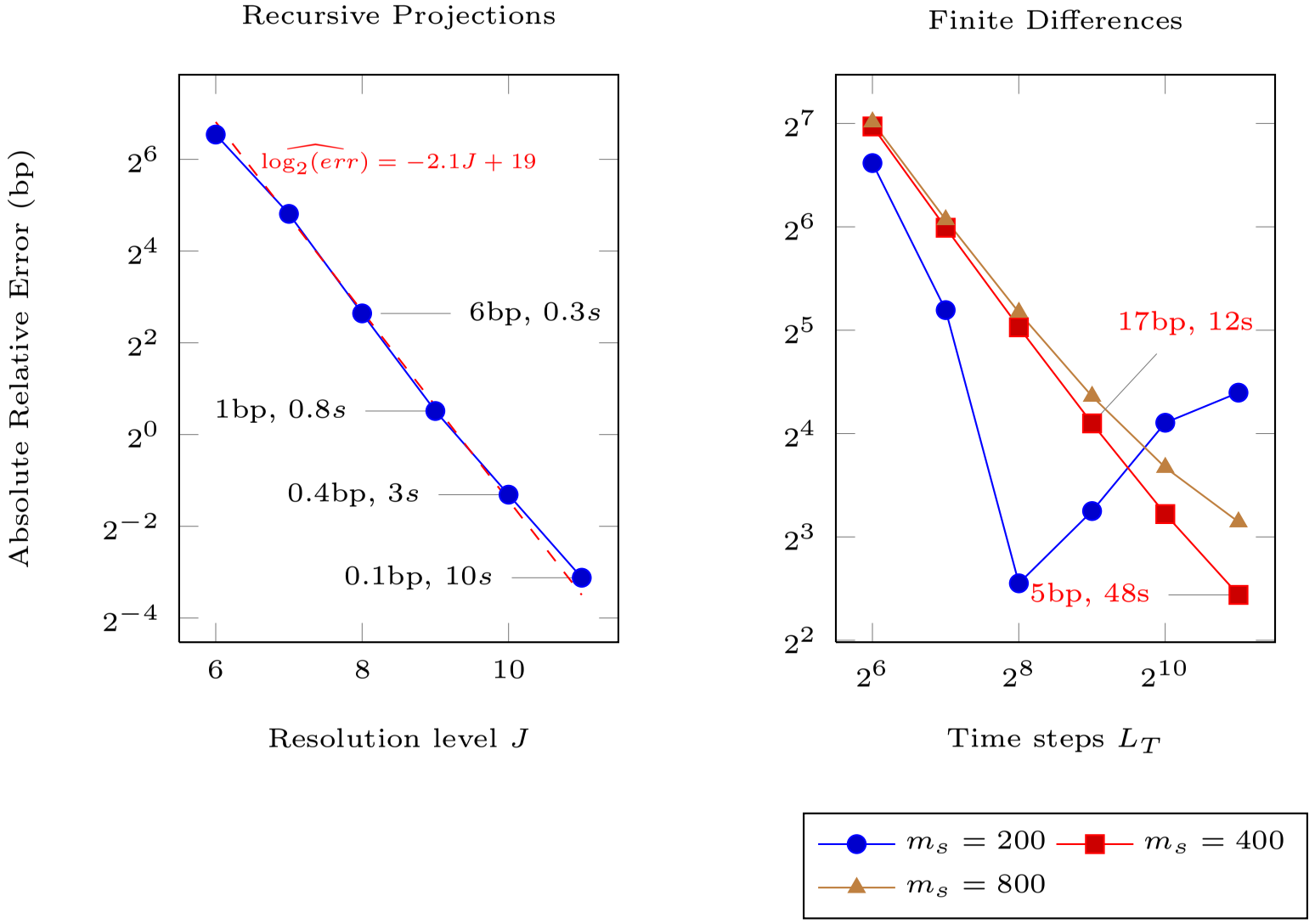

Figure 8: Comparison between the finite difference scheme and the recursive projections on an American call option written on a dividend-paying stock in the Heston case. The option has a maturity of 1 year and a dividend $d = 10$ is paid at $t_l = 0.5$. Other parameters are set equal to $S_0 = 100$, $K = 100$, $r = 0.05$, $\sigma_{LT} = 0.2$, $\beta = 2$, $\omega = 0.2$. The parameter m_s gives the umber of points in the $X_t = \log S_t$ grid for the *FD* scheme, while the σ^2 grid has $m_w = 31$ points. The sampling grid for the $X_t = \log S_t$ variable in the recursive projections has size 2^J. The resolution level in the σ^2 dimension is $J_w = 4$. The dashed line in the left panel corresponds to a fitted linear regression, and shows that the estimated slope is close to the slope of -2 predicted by the theoretical convergence results.

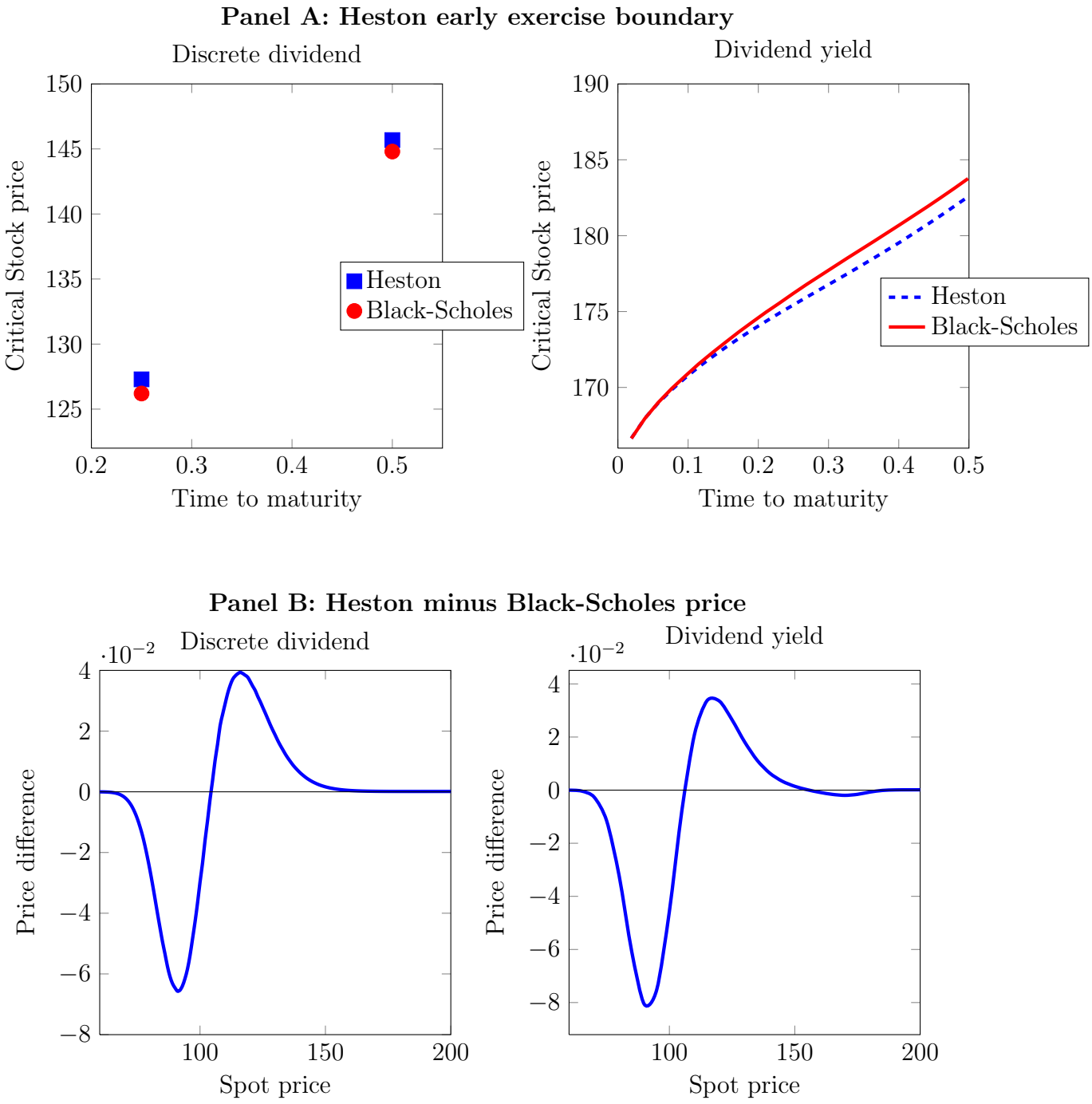

Figure 9: Panel A. Comparison between the early exercise boundary in the Heston and Black-Scholes models of an American call with maturity 6 months, in the case in which the stock pays a dividend yield $r_d = 0.03$ (right) and in the case in which the stock distributes an equivalent quarterly discrete dividend of $d = 1.38$ (left). The remaining parameters are: $K = 100$, $r = 0.05$, $\sigma_0 = 0.2$, $\omega = 0.1$, $\sigma_{LT} = 0.3$, $\beta = 4$, $\rho = -0.5$. We set the volatility parameter in the Black-Scholes model equal to the volatility of the underlying return over the life of the option in the Heston model. Panel B. Heston minus Black-Scholes price of an American call with $T = 0.25$ for different values of S_0 in the case of discrete dividend (left) and continuous dividend yield (right). The remaining parameters are the same as in Panel A.

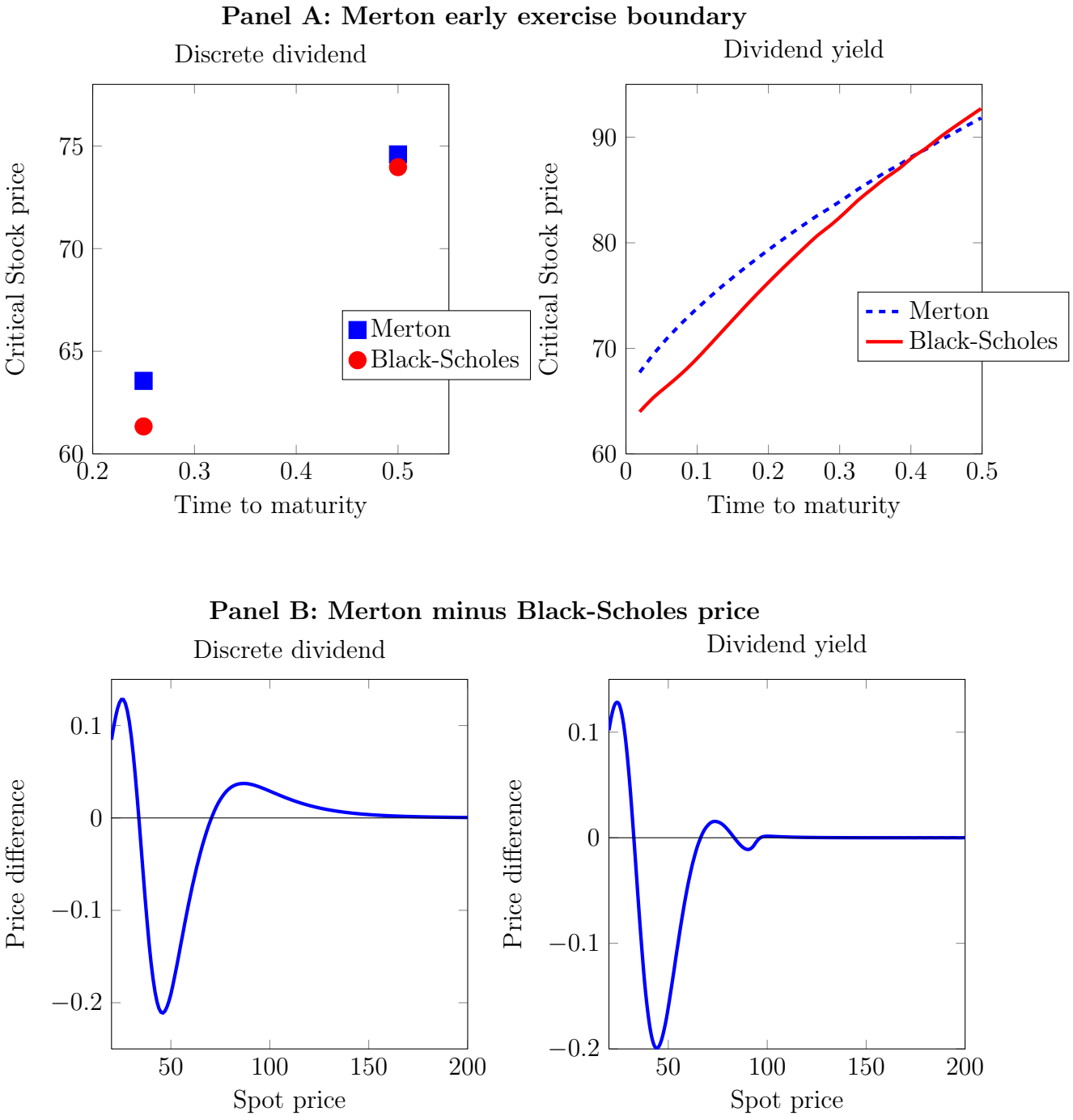

Figure 10: Panel A. Comparison between the early exercise boundary in the Merton and Black-Scholes models of an American call with maturity 6 months, in the case in which the stock pays a dividend yield $r_d = 0.05$ (right) and in the case in which the stock distributes an equivalent quarterly discrete dividend of $d = 1.125$ (left). The other parameters are the following: $K = 40$, $T = 0.5$, $r = 0.08$, $\gamma = 5$, $\sigma_M^2 = 0.05$, $\sigma_\psi^2 = 0.05$, $\mu_\psi = 0$. We set the volatility parameter in the Black-Scholes model equal to the volatility of the underlying return over the life of the option in the Merton model. Panel B. Merton minus Black-Scholes price of an American call with the same parameters as those used in Panel A but different values of S_0 in the case of discrete dividend (left) and continuous dividend yield (right).

			Number of option quotes					
			maturity<60 days			maturity>60 days		
	Ticker	Stock	OTM	ATM	ITM	OTM	ATM	ITM
1	AA	ALCOA INC	42827	6379	41206	28788	9621	29371
2	AXP	AMERICAN EXPRESS CO	43360	10386	59656	27744	14941	50002
3	BAC	BANK OF AMERICA CO	53785	7767	57794	48365	15446	58196
4	BA	BOEING CO	41616	7385	52422	36022	13312	50357
5	CAT	CATERPILLAR INC DEL	40592	8651	55672	34438	15377	54361
6	CHV	CHEVRON CORPORATION	33444	5922	50092	28088	11238	51676
7	CSCO	CISCO SYS INC	58276	8194	63177	44295	14142	55598
8	KO	COCA COLA CO	34710	5055	44407	29054	9249	44421
9	DIS	DISNEY WALT CO	40138	7052	47571	31350	10730	40773
10	XOM	EXXON MOBIL CORP	39509	6427	55468	27944	9425	48375
11	GE	GENERAL ELECTRIC CO	49637	6734	58779	42696	13072	58291
12	HWP	HEWLETT PACKARD CO	54913	9182	53498	47707	17077	51029
13	HD	HOME DEPOT INC	42968	7052	56732	36047	13413	53921
14	INTC	INTEL CORP	55825	8666	61243	43796	14327	52367
15	IBM	INTER. BUS. MACHS	70249	9060	85569	45355	13330	62979
16	JNJ	JOHNSON & JOHNSON	34175	4551	47524	22974	7074	43322
17	MCD	MCDONALDS CORP	33730	5674	47969	27630	10145	48000
18	MRK	MERCK & CO INC	41559	7612	53181	31901	11486	47534
19	MSFT	MICROSOFT CORP	68396	8725	78393	50253	14163	65728
20	MMM	3M CO	36835	6737	46521	24402	10205	38750
21	JPM	MORGAN J P & CO INC	6991	3180	9940	4428	5874	10568
22	PFE	PFIZER INC	51100	6170	54751	47414	12109	52358
23	PG	PROCTER & GAMBLE CO	36971	5904	52782	25422	8974	46956
24	T	AT&T INC	42547	5542	52123	32851	9496	46524
25	TRV	TRAVELERS COMPANIES INC	21404	3803	27783	15069	5811	23646
26	UTX	UNITED TECHNOLOGIES CORP	34765	6366	48200	28973	11712	45580
27	UNH	UNITEDHEALTH GROUP INC	39924	8864	58885	32757	16724	56451
28	VZ	VERIZON COMMUNICATIONS INC	39642	6461	56527	34306	11100	54563
29	WMT	WAL-MART STORES INC	37172	5668	51904	31679	10429	52890
30	DD	DU PONT E I NEMOURS & CO	35262	7404	46302	24077	11097	41173

Table I: Number of observations for in-the-money (ITM), at-the-money (ATM) and out-of-the-money (OTM) call option quotes for the stocks which are the constituents of the Dow Jones Industrial Average Index (DJIA). The data are further broken down by maturity. According to the classification of Bollen and Whaley (2004), a call option is considered OTM if its delta is less then 0.375, ATM if its delta ranges between 0.375 and 0.625 and ITM if its delta is above 0.625.

Underlying	BS	MRT				BTS							
	σ_{BS}	γ	σ_M	σ_ψ	μ_ψ	γ	σ_ψ	μ_ψ	ω	σ_{LT}	β	ρ	σ_0
All stocks	0.29	1.33	0.22	0.16	-0.12	0.50	0.18	-0.12	0.75	0.32	1.52	-0.35	0.28
SP500*	0.18	NA	NA	NA	NA	0.61	0.14	-0.09	0.4	0.2	3.93	-0.52	0.2

Table II: Average values of the parameters of the models of Black-Scholes (BS), Merton (MRT) and Bates (BTS) calibrated at each day before the ex-dividend date on the options written on the dividend-paying stocks belonging to the Dow Jones Industrial Average Index (DJIA). In total we computed 1701 calibrations and the reported values are the averages across these calibrations.
The in-sample sum of squared error is on average equal to 0.26 for the Black-Scholes model, 0.20 for the Merton model, and 0.16 for the Bates model with stochastic volatility.

*Calibrated parameters of the SP500 dynamics are from Bakshi et al. (1997).

	Average rational implied fee			%Implied fee > 0.4446		
Underlying	BS	MRT	BTS	BS	MRT	BTS
All stocks	7.54	7.27	7.23	94%	93%	93%

Table III: The table reports the average implied fee per share which would explain the non-exercise behaviour of investors in each model: Black-Scholes (BS), Merton (MRT) and Bates (BTS). The average implied fee is calculated for each option that should be exercised but which is not optimally exercised by some of the investors as the value of the trading costs $\mathscr{F}$ which makes the continuation value of the option equal to the early exercise proceeds: $C(S-d, K+\mathscr{F}, T) = (S-K-\mathscr{F})$. In the last three columns of the table we report the percentage of options for which the fee that would explain the suboptimal non-exercise behaviour is higher than the conservative fee of 0.4446 dollar per share estimated by Pool et al. (2008).

	Model	Without fee	With fee
Contracts outstanding		406 414 980	
Total market value		99 392 927 000	
Contracts that should be exercised	BS	38 527 586 (9.48%)	31 551 786 (7.76%)
	MRT	30 633 542 (7.54%)	25 340 009 (6.23%)
	BTS	30 486 666 (7.5%)	25 050 616 (6.16%)
Contracts that are left suboptimally non-exercised	BS	15 214 908 (39.49%)	11 077 913 (35.11%)
	MRT	9 404 406 (30.70%)	7 118 002 (28.09%)
	BTS	8 786 524 (28.82%)	6 702 799 (26.76%)
Money available due to exercise opportunity	BS	770 287 766 (0.77%)	647 088 372 (0.65%)
	MRT	555 922 918 (0.56%)	460 980 123 (0.46%)
	BTS	617 173 686 (0.62%)	509 713 158 (0.51%)
Total loss due to suboptimal non-exercise	BS	209 284 628 (27.17%) (0.21%)	172 424 967 (26.65%) (0.17%)
	MRT	133 130 786 (23.95%) (0.13%)	108 951 652 (23.63%) (0.11%)
	BTS	147 480 996 (23.9%) (0.15%)	123 615 413 (24.25%) (0.12%)

Table IV: Summary results of the total loss due to suboptimal non-exercise for the short-term call option series. The numbers are calculated for each series and each day before the ex-dividend date separately, and then pooled together.
The number of contracts outstanding is the total open interest of all contracts two days before the ex-dividend date. The contracts that should be exercised under a specific model, i.e. Black-Scholes (BS), Merton (MRT) and Bates (BTS), are the contracts outstanding for which the continuation value is lower than the exercise proceeds. The number of contracts that are left suboptimally non-exercised is the sum of the open interests one day before ex-dividend of the contracts that should have been exercised. We compute the other quantities in the table in the following way:
Total market value = Contracts outstanding×Market price×100,
Money available = $\max\{0, (S - K - \mathscr{F} - \text{Continuation value}) \times \text{Contracts outstanding} \times 100\}$,
Total loss = $\max\{0, (S - K - \mathscr{F} - \text{Continuation value}) \times \text{Open interest}_{t-1} \times 100\}$,
where $\mathscr{F}$ is the exercise fee. In the first column the results are computed considering $\mathscr{F} = 0$, while in the second column the results are computed considering the conservative fee of 0.44 dollar. The first percentage in parenthesis in the Total loss due to non-exercise is computed with respect to the money available due to exercise opportunities, while the second one is computed with respect to the total market value.

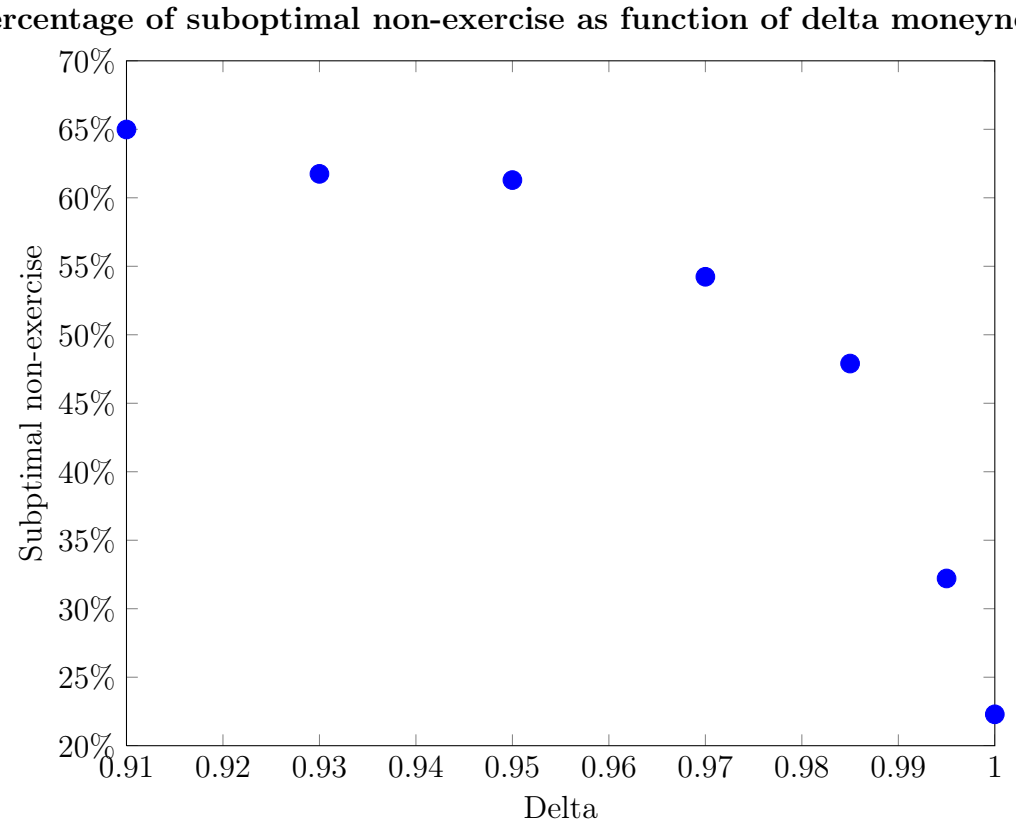

Figure 11: Percentage of contracts suboptimally non-exercised in the model of Black-Scholes as function of the delta moneyness of the contracts.