Early exercise decision in American options with dividend, stochastic volatility and jumps

- ONLINE APPENDIX -

Appendice A provides the proof of Propositions 1 of the main paper. In Appendix B, we provide the analytic form of $\widehat{e}_j(y)$. In Appendix C, we characterize the space translation invariance property of the transition matrices and we describe how we take advantage of this property in the implementation of the algorithm. Appendix D compares the recursive projections method with recent numerical technique which can accommodate discrete dividends, assuming a Black-Scholes dynamics for the underlying security. These methods are the binomial tree and its improved version provided by Vellekoop and Nieuwenhuis (2006), and the simulation least square approach method of Longstaff and Schwartz (2001). We also discuss the new duality approach method of Haugh and Kogan (2004), Rogers (2002), and Andersen and Broadie (2004). Appendix E gives a detailed description of the data and the of calibration procedure, as well as the results of the calibration with a breakdown per stock. Appendix F provides further evidence on the importance of a correct modelling of the dividend as a discrete cash flow when computing the early exercise boundary. In Appendix G we describe the relative advantages of recursive projections and *ADI* schemes in solving different pricing problems.

Appendix A. Proofs

We prove Proposition 1 in three steps. Before stating starting the proof, we start by providing some definitions that we will use extensively.

$$V^{\perp}(y,w,T) = \sum_{j,q} \int d\theta_1 \int d\theta_2 V(\theta_1,\theta_2,T) e_j(\theta_1)\varepsilon_q(\theta_2) e_j(\vartheta)\varepsilon_q(w) \stackrel{\text{def}}{=} \sum_{jq} V^{\perp}_{jq} e_j(y)\varepsilon_q(w), \tag{1}$$

$$\begin{aligned} G_2^{\perp}(y_i,w_p,t;y,w,T) &= \sum_{j,q} \int d\theta_1 \int d\theta_2 G_2(y_i,\xi_p,t;\theta_1,\theta_2,T) e_j(\theta_1)\varepsilon_q(\theta_2) e_j(y)\varepsilon_q(w) \\ &\stackrel{\text{def}}{=} \sum_{j,q} G^{\perp}_{2,ipjq} e_j(y)\varepsilon_q(w). \end{aligned} \tag{2}$$

Equations (1) and (2) define the coefficients $\{V^{\perp}_{jq}\}_{j\in\mathbb{Z},q\in\mathbb{N}}$ and $\{G^{\perp}_{2,ipjq}\}_{j\in\mathbb{Z},q\in\mathbb{N}}$ of the orthogonal projections $V^{\perp}(y,w,T)$ and $G_2^{\perp}(y_i,w_p,t;y,w,T)$. Due to the orthogonality of the orthonormal sets $\{e_j(y)\}_{j\in\mathbb{Z}}$ and $\{\varepsilon_q(y)\}_{q\in\mathbb{N}}$, we obtain the following:

$$v^{\perp}_{ip}(t) \stackrel{\text{def}}{=} \int dy \int dw\, G_2^{\perp}(y_i,w_p,t;y,w,T) V^{\perp}(y,w,T) = \sum_{jq} G^{\perp}_{2,ipjq} V^{\perp}_{jq}. \tag{3}$$

Moreover, we denote the following approximation by $v^{\star}_{ip}(t)$:

$$v^{\star}_{ip}(t) = \sqrt{\Delta y \Delta w} \sum_{jq} \Gamma_2(y_i,w_p,t;y_j,w_q,T) V(y_j,w_q,T). \tag{4}$$

Equation (4) gives approximation of the continuation value at t when the input at time T is a true value. Most of the times, in practical applications, $V(y,w,T) = H(y,T)$. In the last case, $v^{\star}_{ip}(t)$ is the approximation of the price at t of a European contract.

Let $\underline{w}_q$ and $\overline{w}_q$ be the bounds of the support of size Δw of the indicator functions $\{\varepsilon_q\}_{q\in\mathbb{Z}}$ centered on the $\{w_q\}_{q\in\mathbb{Z}}$ grid points. In the following, we repeatedly use the second order Taylor expansion of bivariate functions. Let $\chi(\xi_1,\xi_2)$ be twice differentiable

in the two variables ξ_1 and ξ_2. Then, for $\hat{\xi}_1 \in [\underline{y}_j, \overline{y}_j)$ and $\hat{\xi}_2 \in [\underline{w}_q, \overline{w}_q)$:

$$\chi(\xi_1,\xi_2) = \chi(y_j,w_q) + \partial_{\xi_1}\chi(y_j,w_q)(\xi_1 - y_j) + \partial_{\xi_2}\chi(y_j,w_q)(\xi_2 - w_q)$$
$$+ \partial^2_{\xi_1\xi_2}\chi(\hat{\xi}_1,\hat{\xi}_2)(\xi_1 - y_j)(\xi_2 - w_q) + \frac{1}{2}\partial^2_{\xi_1^2}\chi(\hat{\xi}_1,\hat{\xi}_2)(\xi_1 - y_j)^2 + \frac{1}{2}\partial^2_{\xi_2^2}\chi(\hat{\xi}_1,\hat{\xi}_2)(\xi_2 - w_q)^2.$$

We then have the useful expansion:

$$\begin{aligned}
&\int_{\underline{y}_j}^{\overline{y}_j} d\xi_1 \int_{\underline{w}_q}^{\overline{w}_q} d\xi_2\, \chi(\xi_1,\xi_2) \\
&= \chi(y_j,w_q)\Delta y \Delta w + \frac{1}{2}\int_{\underline{y}_j}^{\overline{y}_j} d\xi_1 \int_{\underline{w}_q}^{\overline{w}_q} d\xi_2 \Big[\partial^2_{\xi_1^2}\chi(\hat{\xi}_1,\hat{\xi}_2)(\xi_1 - y_j)^2 + \partial^2_{\xi_2^2}\chi(\hat{\xi}_1,\hat{\xi}_2)(\xi_2 - w_q)^2\Big] \\
&= \chi(y_j,w_q)\Delta y \Delta w + O\big(\underline{\Delta}^3\big), \qquad as\ \underline{\Delta} \to 0, \qquad (5)
\end{aligned}$$

because, since y_j and w_q are the centre points of the integration interval, the integrals of the other terms of the expansion vanish.

The proof of Proposition 1 is organized in the following four steps: 1) Lemma 1 tells us that what matters for the convergence properties in Equations (14) and (15) of the main paper is the rate of convergence of the approximated continuation value to the true continuation value. 2) In Lemma 2, we show that the computed continuation value $v^{\star}_{ip}(t)$ verifies $v^{\star}_{ip}(t) = v^{\perp}_{ip}(t) + O(\underline{\Delta}^2)$. 3) In Lemma 3, we prove that $v^{\perp}_{ip}(t) = V(y_i, w_p, t) + O(\underline{\Delta}^2)$, which entails that $v^{\star}_{ip}(t) = V(y_i, w_p, t) + O(\underline{\Delta}^2)$, which proves Proposition 1 in the European option case, i.e. Equation (14) by setting $t = t_{L-1}$. 4) In Lemma 4, we conclude by proving the recursive formula of Equation (15) of the main paper. The summations on the indices j and q are understood to be from $-\infty$ to $+\infty$ and from 1 to $+\infty$, respectively.

LEMMA 1. Let A_1, A_2 and a_2 be real numbers such that A_1 and A_2 are true quantities and a_2 is an approximation of A_2. Then, we have the following inequality:

$$|\max\{A_1, a_2\} - \max\{A_1, A_2\}| \leq |a_2 - A_2|. \qquad (6)$$

Inequality (6) shows that the rate of convergence of $\max\{A_1, a_2\}$ to $\max\{A_1, A_2\}$ is given by the rate of convergence of a_2 to A_2.

Proof of Lemma 1:

Proof. We must analyze four possibilities.

1. if $A_1 > a_2$ and $A_1 > A_2$, then $|\max\{A_1, a_2\} - \max\{A_1, A_2\}| = 0$.
2. if $A_1 \leq a_2$ and $A_1 \leq A_2$, then $|\max\{A_1, a_2\} - \max\{A_1, A_2\}| = |a_2 - A_2|$.
3. if $A_1 > a_2$ and $A_1 \leq A_2$, then we have $|\max\{A_1, a_2\} - \max\{A_1, A_2\}| = |A_1 - A_2| \leq |a_2 - A_2|$, because A_1 lies between a_2 and A_2.
4. if $A_1 \leq a_2$ and $A_1 > A_2$, then we have $|\max\{A_1, a_2\} - \max\{A_1, A_2\}| = |a_2 - A_1| \leq |a_2 - A_2|$, because A_1 lies between a_2 and A_2.

Gathering points 1-4 yields inequality (6). □

LEMMA 2. The approximation error between $v_{ip}^{\perp}(t)$ and $v_{iq}^{\star}(t)$ defined in (3) and (4) satisfies:

$$v_{ip}^{\star}(t) = v_{ip}^{\perp}(t) + O(\underline{\Delta}^2), \qquad as\ \underline{\Delta} \to 0.$$

Proof of Lemma 2. We must bound the difference:

$$\sum_{jq} \left| \sqrt{\Delta y \Delta w}\ \Gamma_2(y_i, w_p, t; y_j, w_q, T) V(y_j, w_q, T) - G_{2,ipjq}^{\perp} V_{jq}^{\perp} \right|.$$

By Fourier isometry, we have:

$$\iint d\theta_1 d\theta_2 G_2(y_i, w_p, t; \theta_1, \theta_2, T) e_j(\theta_1) \varepsilon_q(\theta_2) = \frac{1}{4\pi^2} \iint d\lambda d\kappa \hat{G}_2(y_i, w_p, t; \lambda, \kappa, T) \hat{e}_j(-\lambda) \hat{\varepsilon}_q(-\kappa).$$

Then, we deduce:

$$\sum_{jq}\Big|\sqrt{\Delta y\Delta w}\;\Gamma_2(y_i,w_p,t;y_j,w_q,T)V(y_j,w_q,T)-G^{\perp}_{2,ipjq}V^{\perp}_{jq}\Big|$$
$$=\frac{1}{4\pi^2}\sum_{jq}\Big|\sqrt{\Delta y\Delta w}\Big(\sum_{r,z=-\infty}^{\infty}\hat{G}_2(x_i,w_p,t;\lambda_r,\kappa_z,T)\hat{e}_j(-\lambda_r)\hat{\varepsilon}_q(-\kappa_z)\Delta\lambda\Delta\kappa\Big)V(y_j,w_q,T) \tag{7}$$
$$-\Big(\iint d\lambda d\kappa\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)\hat{e}_j(-\lambda)\hat{\varepsilon}_q(-\kappa)\Big)\Big(\iint d\theta_1 d\theta_2 V(\theta_1,\theta_2,T)e_j(\theta_1)\varepsilon_q(\theta_2)\Big)\Big|.$$

The functions $\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)$, $\hat{e}_j(-\kappa)$ and $\hat{\varepsilon}_j(-\lambda)$ are twice continuously differentiable. Moreover, let $\overline{\Delta}=\sqrt{\Delta\kappa^2+\Delta\lambda^2}$. Using the property (5) with $\chi(\lambda,\kappa)=\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)\hat{e}_j(-\lambda)\hat{\varepsilon}_q(-\kappa)$ we have that:

$$\iint d\lambda d\kappa\,\hat{G}_2(y_i,w_p,t;\lambda,\kappa,T)\hat{e}_j(-\lambda)\hat{\varepsilon}_q(-\kappa)+O(\overline{\Delta}^3), \tag{8}$$
$$=\sum_{r,z=-\infty}^{\infty}\hat{G}_2(x_i,\xi_p,t;\lambda_r,\kappa_z,T)\hat{e}_j(-\lambda_r)\hat{\varepsilon}_q(-\kappa_z)\Delta\lambda\Delta\kappa \qquad \text{as } \overline{\Delta}\to 0.$$

Exploiting the continuity property of $V(y,w,T)$, we obtain:

$$\Big|\sqrt{\Delta y\Delta w}\iint d\theta_1 d\theta_2 V(\theta_1,\theta_2,T)e_j(\theta_1)\varepsilon_q(\theta_2)-\;\Delta y\Delta w\,V(y_j,w_q,T)\Big| \tag{9}$$
$$\le\int_{\underline{y}_j}^{\overline{y}_j}d\theta_2\int_{\underline{w}_q}^{\overline{w}_q}d\theta_1\Big|V(\theta_1,\theta_2,T)-V(y_j,w_q,T)\Big|\le\Delta y\Delta w\,C\underline{\Delta}=O(\underline{\Delta}^2),\qquad\text{as }\underline{\Delta}\to 0.$$

It then suffices to substitute Equation (9) and (8) into (7), and to choose $\overline{\Delta}=O(\underline{\Delta})$, to prove the statement of Lemma 2.

LEMMA 3. The following equality holds:

$$v^{\perp}_{ip}(t)=V(y_i,w_p,t)+O(\underline{\Delta}^2),\qquad as\ \underline{\Delta}\to 0.$$

Proof of Lemma 3. We study the following difference:

$$V(y_i,w_p,t)-v_{ip}^{\perp}(t)=\iint dydwG_2(y_i,w_p,t;y,w,T)V(y,w,T)-\sum_{jq}G_{2,ipjq}^{\perp}V_{jq}^{\perp} \tag{10}$$

$$=\sum_{jq}\int_{\underline{y}_j}^{\overline{y}_j}dy\int_{\underline{w}_q}^{\overline{w}_q}dw\,G_2(y_i,w_p,t;y,w,T)V(y,w,T)$$

$$-\sum_{jq}\frac{1}{\Delta y\Delta w}\int_{\underline{y}_j}^{\overline{y}_j}d\theta_1\int_{\underline{w}_q}^{\overline{w}_q}d\theta_2 G_2(y_i,w_p,t;\theta_1,\theta_2,T)\int_{\underline{y}_j}^{\overline{y}_j}d\vartheta_1\int_{\underline{w}_q}^{\overline{w}_q}d\vartheta_2 V(\vartheta_1,\vartheta_2,T).$$

We show that both terms in the right-hand side of (10) are equal to $\sum_{j,q}G_2(y_i,w_p,t;y_j,w_q,T)V(y_j,w_q,T)\Delta y\Delta w+O(\underline{\Delta}^2)$. We start by the generic term of the first summation. By applying (5) and (9):

$$\Big|\int_{\underline{y}_j}^{\overline{y}_j}dy\int_{\underline{w}_q}^{\overline{w}_q}dw\,G_2(y_i,w_p,t;y,w,T)V(y,w,T)-G_2(y_i,w_p,t;y_j,w_q,T)V(y_j,w_q,T)\Big|\leq$$

$$\int_{\underline{y}_j}^{\overline{y}_j}dy\int_{\underline{w}_q}^{\overline{w}_q}dw\,G_2(y_i,w_p,t;y,w,T)\big|V(y,w,T)-V(y_j,w_q,T)\big|$$

$$+V(y_j,w_q,T)\int_{\underline{y}_j}^{\overline{y}_j}dy\int_{\underline{w}_q}^{\overline{w}_q}dw\,\big|G_2(y_i,w_p,t;y,w,T)-G_2(y_i,w_p,t;y_j,w_q,T)\big|$$

$$\leq\sup_{\substack{y\in[\underline{y}_j,\overline{y}_j)\\w\in[\underline{w}_q,\overline{w}_q)}}G_2(y_i,w_p,t;y,w,T)O\big(\underline{\Delta}^2\big)+\sup_{\substack{y\in[\underline{y}_j,\overline{y}_j)\\w\in[\underline{w}_q,\overline{w}_q)\\\xi\in\{y,w\}}}\partial_{\xi^2}^2G_2(y_i,w_p,t;y,w,T)O\big(\underline{\Delta}^3\big).$$

We can easily check that the generic term of the second summation in Equation (10) equals:

$$\frac{1}{\Delta y\Delta w}\Big[G_2(y_i,w_p,t;y_j,w_q,T)\Delta y\Delta w+O(\underline{\Delta}^3)\Big]\Big[V(y_j,w_q,T)\Delta y\Delta w+O(\underline{\Delta}^2)\Big]$$

$$=G_2(y_i,w_p,t;y_j,w_q,T)V(y_j,w_q,T)\Delta y\Delta w+G_2(y_i,w_p,t;y_j,w_q,T)O(\underline{\Delta}^2).$$

Then:

$$|V(y_i,w_p,t)-v_{ip}^{\perp}(t)|\leq\sum_{j,q}\sup_{\substack{y\in[\underline{y}_j,\overline{y}_j)\\w\in[\underline{w}_q,\overline{w}_q)}}G_2(y_i,w_p,t;y,w,T)O\big(\underline{\Delta}^2\big).$$

Because $G_2(y_i,w_p,t;y,w,T)$ is a density, the above summation is finite, and it proves Lemma 3. Combining the results of Lemma 2 and Lemma 3, we have shown that the

continuation value $v^{\star}_{ip}(t) = V(y_i, w_p, t) + O(\underline{\Delta}^2)$, that is, we have proven the convergence of the algorithm in the European case.

LEMMA 4. Let $v_{ip}(t_l)$ be defined as in Equation (16) of the main paper, with $l = 1, \ldots, L-2$. Then $v_{ip}(t_l)$ converges to the true price $V(y_i, w_p, t_l)$ at a rate of the order $O(\underline{\Delta}^2)$.

Proof. We start by showing the convergence of $v_{ip}(t_{L-2})$ to $V(y_i, w_p, t_{L-2})$. Because of Lemma 1, we only need to prove the convergence of the approximated continuation value at $t = t_{L-2}$ to the true continuation value at $t = t_{L-2}$. We consider a contract evaluated at two dates $\{t_{L-2}, t_{L-1}\}$ prior to maturity, $t_L = T$, namely $t_{L-2} < t_{L-1} < T$. Then:

$$
\begin{aligned}
&\sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) v_{jq}(t_{L-1}) \sqrt{\Delta y \Delta w} \\
&= \sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) \big[V(y_j, w_q, t_{L-1}) - V(y_j, w_q, t_{L-1}) + v_{jq}(t_{L-1}) \big] \sqrt{\Delta y \Delta w} \\
&= \sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) V(y_j, w_q, t_{L-1}) \sqrt{\Delta y \Delta w} \\
&\quad + \sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) \big[v_{jq}(t_{L-1}) - V(y_j, w_q, t_{L-1}) \big] \sqrt{\Delta y \Delta w}.
\end{aligned}
$$

The quantities $\{V(y_j, w_q, t_{L-1})\}_{j \in \mathbb{Z}, p \in \mathbb{N}}$ are exact values; thus it follows from Lemma 2 and Lemma 3 that:

$$
\sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) V(y_j, w_q, t_{L-1}) \sqrt{\Delta y \Delta w} = V(y_i, w_q, t_{L-2}) + O\big(\underline{\Delta}^2\big).
$$

Again, from Lemmas 1 and 3, it follows that $v_{jq}(t_{L-1}) = \max\{v^{\star}_{jq}(t_{L-1}), H(y_j, w_q, t_{L-1})\} = V(y_j, t_{L-1}) + O\big(\underline{\Delta}^2\big)$. Then:

$$
\begin{aligned}
&\sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) \big[v_{jq}(t_{L-1}) - V(y_j, w_q, t_{L-1}) \big] \sqrt{\Delta y \Delta w} \\
&\leq \sup_j |v_{jq}(t_{L-1}) - V(y_j, w_q, t_{L-1})| e^{-r(t_{L-1} - t_{L-2})} \big(1 + O\big(\underline{\Delta}^3\big)\big) = O\big(\underline{\Delta}^2\big).
\end{aligned}
$$

In the last inequality, we take advantage of the fact that

$$\sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) \sqrt{\Delta y \Delta w} = e^{-r(t_{L-2}-t_{L-1})}\big(1 + O\big(\underline{\Delta}^3\big)\big),$$

because $G(x, \xi, t_{L-2}; y, w, t_{L-1})$ is the deterministic discount factor times a density. Indeed, the approximation operators built on indicator functions are shape preserving, (see Dechevsky and Penev (1997) and Cosma et al. (2007)), and the property of integration to one of a density is preserved. The $O\big(\underline{\Delta}^3\big)$ term is the speed at which the sum $\sum_{jq} \Gamma(y_i, w_p, t_{L-2}; y_j, w_q, t_{L-1}) \sqrt{\Delta y \Delta w}$ converges to $\int dydw \sum_{jq} \int d\theta_1 \theta_2 G(y_i, w_p, t_{L-2}; \theta_1, \theta_2, T_{L-1}) e_j(\theta_1) \varepsilon_q(\theta_2) e_j(y) \varepsilon_q(w)$, and can be checked using the same series expansions techniques as in the proof of Lemma 2. It readily follows that $v_i(t_{L-2}) = V(y_i, t_{L-2}) + O\big((\Delta y)^2\big)$. The extension to prior dates $t_l = t_{L-3}, t_l = t_{L-4}, \ldots$, immediately follows by recursively applying the same arguments used above.

□

The proof of Proposition 1 can be performed in a more general framework, and for basis sets other than indicator functions. The key requirement is that only a finite number of basis functions contribute to the the approximation of a function at a given point (y_i, w_p). Examples are orthonormal wavelets, non-orthogonal and bi-orthogonal wavelet bases, and B-splines. The use of these function bases may be useful when we need a basis that better adapts to the specific geometry of more complicated pricing problems.

Appendix B. Analytic Form for $\hat{e}_j(y)$

Let $e_j(y)$ be the indicator function of the interval $[\underline{y}_j, \overline{y}_j)$, normalized according to the L_2 norm. Then the Fourier transform $\hat{e}_j(\lambda)$ is given by the following:

$$\begin{aligned}\hat{e}_j(\lambda) &= \frac{2}{\sqrt{\overline{y}_j - \underline{y}_j}}\left[\cos\left(\frac{\overline{y}_j + \underline{y}_j}{2}\lambda\right) + \iota \sin\left(\frac{\overline{y}_j + \underline{y}_j}{2}\lambda\right)\right]\sin\left(\frac{\overline{y}_j - \underline{y}_j}{2}\lambda\right)\Big/\lambda \\ &= \frac{2}{\lambda\sqrt{\Delta y}}\left[\cos(y_j\lambda) + \iota\sin(y_j\lambda)\right]\frac{\sin\left(\Delta y\lambda\right)}{\lambda}.\end{aligned}$$

Appendix C. Space Translation Invariance Property of Transition Matrices

The transition matrix $\mathbf{\Gamma}_2(y_i, w_p, t_l; t_{l+1})$, as defined in Equation (11) of the main paper, is a function of the conditioning values (y_i, w_p). The following two remarks greatly simplify and speed up the computation of the transition matrices. First, in Equation (17) of the main paper, only $\hat{\mathbf{G}}_2(y_i, w_p, t_l; t_{l+1})$ actually depends on y_i and w_p, which means that $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ only need to be computed once. Second, the evolution of the asset prices logarithm in the stochastic volatility model has the property that increments are independent of the price level. Let $M_2(\log(x), \xi, t_l; \log(y), w, t_{l+1}) = G_2(x, \xi, t_l; y, w, t_{l+1})y$ be the bivariate state price density as a function of $\log(y)$ and let $\hat{M}_2(\log(x), \xi, t_l; \lambda, \kappa, t_{l+1})$ be its Fourier transform. Then, if we sample the transition densities and the payoff function on an equispaced grid in the $\log(y)$ scale, i.e. $\{\log(y)_i\}_{i=1,\ldots,N}$, we can rewrite (17) by defining the $N \times W$ matrix $\mathbf{\Psi}_2((\log(y))_i, w_p, t_l; t_{l+1}) = \boldsymbol{\phi}'\hat{\mathbf{M}}_2((\log(y))_i, w_p, t_l; t_{l+1})\boldsymbol{\varphi}\sqrt{\Delta y \Delta w}$, where the $R \times Z$ matrix $\hat{\mathbf{M}}_2((\log(y))_i, w_p, t_l; t_{l+1})$ corresponds to the $R \times Z$ matrix $\hat{\mathbf{G}}_2(y_i, w_p, t_l; t_{l+1})$. Equation (18) becomes: $v_{ip}(t_l) = \max\{H(e^{(\log(y))_i}, t_l),\ \mathbf{\Psi}_2((\log(y))_i, w_p, t_l; t_{l+1}) : \boldsymbol{v}_2(t_{l+1})\}$, where $v_{ip}(t_l)$ is now the approximation to the value $V(e^{(\log(y))_i}, w_p, t_l)$. We have that

$\mathbf{\Psi}_{2,jq}((\log(y))_{i+\zeta}, w_p, t_l; t_{l+1}) = \mathbf{\Psi}_{2,j-\zeta\, q}((\log(y))_i, w_p, t_l; t_{l+1})$ for $\zeta \in \mathbb{Z}$, provided that $0 < i+\zeta < N$. We refer to this property as to the space translation invariance property of transition matrices. In implementations, we compute $\mathbf{\Psi}_2((\log(y))_i, w_p, t_l; t_{l+1})$ only once for at-the-money values of $((\log(y))_i, w_p)$, and reconstruct the other transition matrices exploiting the space translation invariance property. Again, this feature exemplifies the computational advantage of direct sampling based on equally-spaced grids.

If we have to take into account discrete dividends, as in Section 3.3 of the main paper, at each dividend date t_h, we must compute the continuation value of the option at the grid $\{(\log(e^{\log(y)_i} - d), w_p)\}_{i=1,\ldots,N;p=1,\ldots,W}$. If the original grid $\{(\log(y)_i, w_p)\}_{i=1,\ldots,N;p=1,\ldots,W}$ has a regular step in the $\log(y)_i$ direction, then this is no more true for the grid $\{(\log(e^{\log(y)_i} - d), w_p)\}_{i=1,\ldots,N;p=1,\ldots,W}$. We can still take advantage of the space translation invariance of the transition matrices because the state price density $M_2(\log(x), \sigma_t^2, t_h; \log(y), w, t_{h+1})$ is a function of $\log(x)$ and $\log(y)$ only through the difference $\log(y) - \log(x)$. Let us perform the following change of variable:

$$
\begin{aligned}
V(x-d, \sigma_t^2, t_h) &= \iint d\log(y) dw M_2(\log(x-d), \sigma_t^2, t_h; \log(y), w, t_{h+1}) V(e^{\log(y)}, w, t_{h+1}) \\
&= \iint d\log(y) dw M_2\Big(\log(x), \sigma_t^2, t_h; \log(y) + \log\big(\frac{x}{x-d}\big), w, t_{h+1}\Big) V(e^{\log(y)}, w, t_{h+1}) \\
&= \iint d\log(y) dw M_2\Big(\log(x), \sigma_t^2, t_h; \log(y), w, t_{h+1}\Big) V(e^{(\log(y)+\log(1-d/x))}, w, t_{h+1}).
\end{aligned}
$$

For pricing by recursive projection, this procedure translates into the relationship: $v_{ip}(t_h) = \max\Big\{H(e^{\log(y)_i}, t_h), \mathbf{\Psi}_2(\log(y)_i, w_p, t_h; t_{h+1}) \,:\, \tilde{\boldsymbol{v}}_{2;d}(t_{h+1})\Big\}$, where $\tilde{\boldsymbol{v}}_{2;d}(t_{h+1})$ are approximations of the value function $V(e^{(\log(y)_j + \log(1-d/e^{\log(y)_i}))}, w_q, t_{h+1})$ obtained by a second-order interpolation of the elements of $\boldsymbol{v}_2(t_{h+1})$. We can still compute the $\mathbf{\Psi}_2(\log(y)_i, w_p, t_h; t_{h+1})$ matrices on the regular grid $\{(\log(y)_i, w_p)\}_{i=1,\ldots,N;p=1,\ldots,W}$, and we can still use the space translation invariance property to speed up computations.

Appendix D. Comparison with Other Methods

As a first numerical example in the Black-Scholes framework, we compare the convergence speed of a binomial tree and of the recursive projections method in pricing an American call option on a dividend-paying stock[1]. Two popular modeling choices for the dividend payment 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. 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. The work of Vellekoop and Nieuwenhuis (2006) provides a recent enhancement of the classical binomial tree method which incorporates discrete dividend payments through an approximation of the continuation value of the option at the ex-dividend dates. This new algorithm has been proven to be substantially faster than the standard non-recombining binomial tree, and is therefore a reliable benchmark for this simulation exercise.

[Figures 1 and 2 about here]

Figure 1 compares the convergence speed of the enhanced binomial tree and that of the recursive projections 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 an increase of speed of approximately a factor 10 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 projections

[1] All of the codes are written in C++. The codes are available from the authors upon request.

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 an additional benchmark, Figure 2 displays the convergence speed of the recursive projections jointly with the one of a standard non-recombining tree. Even though the non-recombining tree is known to be an inefficient method, it is still used as a common reference point in the literature, and we show this graph for comparison purposes. We can see that the gain of speed of the recursive projection is of the order of 10^4. As an aside, for $S_0 = 100$, if we approximate the known constant dividend $d = 2$ with a known continuous dividend yield[2] $r_d = 0.013$, then a binomial tree with 10000 steps delivers a value of 18.213 instead of 18.526, with a relative error of approximately $169bp$. 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. 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. The discontinuity may introduce noise at most in the coefficient relative to the indicator function of the interval in which the discontinuity is

[2]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.

located. The noise is completely eliminated if we make sure that that the strike value lies in between two consecutive grid points, so that the discontinuity is at the boundary between two consecutive indicator functions. In this numerical example, we use the standard binomial tree as a benchmark, since the method of Vellekoop and Nieuwenhuis (2006) provides no advantage in the absence of dividends. Figure 3 (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 3 about here]

Another group of numerical methods that can be applied to the same pricing problems are the Monte-Carlo simulation methods. They can handle both discrete dividends and multidimensional settings. The least-squares approach of Longstaff and Schwartz (LS) provides a simulation based algorithm to price American options, via a lower bound for the true price. This lower bound is then coupled with an upper bound in the implementation of Andersen and Broadie (2004) of the duality approach of Haugh and Kogan (2004) and Rogers (2002). In their numerical results, Andersen and Broadie (2004) show that the gap between the lower bound and the upper bound can be very tight, making the algorithm appealing. In Figure 4, we compare the speed and accuracy of the LS algorithm with our method in the same three examples as before. Our algorithm is faster than the LS method by at least four orders of magnitude. Intuitively, the main advantage of our algorithm is that it needs to evaluate the option only when it can be optimal to exercise

it. In the case of a call option, this happens just before the payment of the dividends. In this specific example, when there are only two dividend payments, our algorithm computes the final price with only two recursions. On the contrary, every simulation based method needs to simulate the entire trajectories, increasing the computation time. The duality approach implementation of Andersen and Broadie (2004) builds on the LS algorithm and necessitates additional simulations at each potential exercise date to build the upper bound for the price, thus further increasing the numerical complexity and the computation time. Given the results obtained for the LS algorithm, we can confidently conclude that our algorithm is also faster than the duality approach in pricing a call option written on a stock which distributes regular discrete dividends.

Appendix E. Data and Calibration Procedure

We conduct our analysis over the period January 1996 - December 2012. We use all short term call option series with maturity less than six months written on the dividend-paying stocks belonging to the Dow Jones Industrial Average Index (DJIA) at the end of 2012. According to other studies (Barraclough and Whaley (2012); Pool et al. (2008)), we proxy for the timing of the expected dividends paid during an option life time with the actual distribution time of dividends, and we proxy for future dividends amounts by using the last known dividend amount. We exclude from our sample the data relative to periods in which the underlying stock experiences an unusual corporate event that may alter the option valuation, such as special dividend distribution, new equity issue, or spin off. To ensure the exclusion of the effects of such corporate events and their anticipation by investors from our sample, we do not consider the data starting from nine months before the special corporate event up to nine months after. After applying these exclusionary criteria, we obtain a dataset of 1701 days before ex-dividend to analyze.

For each stock and at each day before an ex-dividend date, we separately calibrate the

parameters of the models of Black-Scholes, Merton, and Bates on a calibration sample by minimizing the implied volatility mean squared error (IVMSE) as $IVMSE(\chi) = \sum_{i=1}^{n}(\sigma_i - \sigma_i(\chi))^2$, where χ is the set of parameters to estimate, $\sigma_i = BS^{-1}(C_i, T_i, K_i, S, r)$ is the market implied volatility and $\sigma_i(\chi) = BS^{-1}(C_i(\chi), T_i, K_i, S, r)$ is the model implied volatility, where $C_i(\chi)$ is the model price of the American option i. The choice of this loss function follows the argumentation of Christoffersen and Jacobs (2004). The calibration made on implied volatilities is more stable out of sample, in particular for the stochastic volatility model. We infer the model specific parameters of the underlying process by calibration on a set of reliable and liquid option data. More specifically, the calibration sample consists of contracts traded in the four months preceding the calibration day, that have no dividend payment in their remaining life. These contracts can be treated as European. In this way we can take advantage of the semi-closed pricing formula for European options. We then apply some exclusionary criteria: we do not consider options that should be optimally exercised, because their price is equal to the exercise proceeds for mostly all values of the parameters and the minimization problem is ill posed. We thus consider the option quotations that strictly satisfy the following inequality: $C > S - K$; we do not consider option data with a price less than 3/8 of a dollar, in order to avoid effects due to price discreteness; we do not consider options with volume equal to zero as the non-traded quoted prices are not reliable prices; finally, we do not consider options which are deep in-the-money or deep out-of-the-money, as they can destabilize the minimization problem. Following Bollen and Whaley (2004), a call option is classified deep in-the-money if its delta is larger than 0.875. Symmetrically, a call option is classified deep out-of-the-money if its delta is less than 0.125. After applying these criteria, the calibration sample at each day before the ex-dividend date consists of 110 call options on average.

The models of Black-Scholes and Merton are one-dimensional and do not present any particular numerical issue; so we simply calibrate all their parameters on the calibration

sample described above. The Bates model, on the contrary, is two-dimensional. Therefore, in addition to the calibration of its parameters, it needs the calculation of the daily instantaneous spot volatility σ_0, which is a non-observable variable. It is also a more sophisticated model with its seven parameters. In order to efficiently calibrate it, we use a procedure where we take into consideration the specific role of the parameters on the implied volatility surface. To the best of our knowledge, we are the first academic work in which the Bates model is calibrated on single stocks. Hence we borrow some intuition for our new calibration procedure from the practitioners studies of Hagan et al. (2002), and West (2005). According to these studies, each parameter of the volatility dynamics has a specific impact on the term structure of the implied volatility smile. The volatility of volatility ω rules the convexity of the smile while the correlation parameters ρ rules the slope of the smile. In the two works cited above, the authors consider a pure diffusive process for the volatility, ignoring the mean reversion part. They show that for short term options this reduced model provides a very good fit to the data. Indeed, for the short term options the value of Δt is very small compared to the possible values of ΔW and the dynamics of the stochastic volatility process is driven mainly by the Brownian motion part. The role of the mean reverting part of the stochastic volatility process of Heston is to reproduce that implied volatilities of long maturity options are less volatile than those of short maturity options and are usually closer to the long run average volatility. The mean reverting part avoids that the volatility increases indefinitely with maturity. In principle, for our application on short term options, only the diffusion component of the stochastic volatility is sufficient to give a good fit. However, as we want to employ the full dynamics of the Bates model, we calibrate the mean reversion and the long term volatility parameters as well on long term options.

To this end, we calibrate the parameters in two steps: first, we calibrate the jump parameters together with the volatility of volatility and the correlation on the short term

options calibration sample described above. In this optimization, we do not consider the mean reversion part of the stochastic volatility. Then, as a second step, we calibrate the mean reversion and the long term volatility on a sample of two long term options, while keeping the other parameters fixed to those obtained in the previous step. For the calibration, we use the two long term options with the highest trading volume among the long term options with maturity between ten months and two years which were recorded in the four months before the calibration. In this long term calibration, we use as objective function the minimisation of the price percentage mean squared error instead of the implied volatility mean squared error. As the long term options have dividends during their life and their American price differs from the European one, we cannot recover the implied volatility in the usual way. If we had calibrated the long term volatility and the mean reversion on the short term options directly, we would have obtained an unreasonable high value for the mean reversion and an unreasonable low value for the long term volatility. This spurious effect is due to the very high convexity of the short term smile combined with the drift part of the stochastic volatility dynamics having little or null impact on the prices of short term options, as explained before.

For the calculation of the non-observable daily instantaneous spot volatility $\sigma_0(t)$, we follow the result of Medvedev and Scaillet (2010). We use the time series of the one month (or close to) European at-the-money implied volatility as proxy for the spot volatility. For the days considered in the calibration sample we have European options by construction. In addition to these days, we need as well to compute the value of σ_0 on all days before the ex-dividend dates in order to price options and determine which options should be exercised. On the day before the ex-dividend date, however, all options have a dividend during their life, as the first dividend is paid the day after. In principle there are no European options available. In order to make it possible to calculate σ_0 on the day before the ex-dividend date, we consider as European the options which should not

be exercised and which have only the dividend paid the following day left during their remaining life. These options are not European only because they have an early exercise possibility until the dividend is paid. However, as the dividend is paid the following day and these contracts are outside the early exercise region, the early exercise premium is nearly zero and the price of the American option almost coincides with the price of an European option. In the same spirit, Bakshi et al. (2003) extract the European implied volatility from the American options prices, and they show that the difference between the European implied volatility and the American implied volatility is negligible and within the bid-ask spread. By employing the same approximation, we calculate σ_0 at each day before the ex-dividend date as the average of the European implied volatility of the at-the-money options that should not be exercised with maturity one month (or close to).

The results of the calibration with a breakdown per stock are presented in Table 1. We note that the calibrated values of the parameters are homogeneous among stocks, and take sensible values in line with other studies made on index options (see Bakshi et al. (1997)).

Appendix F. Correct Modelling of Dividend

In this section, we provide further evidence on the importance of a correct modelling of the dividend as a discrete cash flow when computing the early exercise boundary.

The setup is the same as in Section 4.1 of the main text. There are two dividends to be paid over the remaining life of the option, which is 6 months. One is to be paid immediately ($t = 0$), the second in 3 months ($t = 0.25$). We compute the early exercise boundary by i) correctly taking into account that the dividends are discrete, and ii) by using the "escrow dividend" approximation, that is we price the contract at $t = 0$ as if it were a European option, with the continuation value after the first dividend payment

not being computed at $S_0 - d$, but at $S_0 - d - d \cdot e^{-r(0.25)}$. We compute the early exercise boundary under the three models (Black-Scholes, Merton and Bates), with the following representative parameters: $\lambda = 5$, $\mu_j = 0$, $\sigma_j = 0.2$, $\sigma_{MRT} = 0.2$, σ_{BS}, $\sigma_0 = 0.2$, $\omega = 0.1$, $\sigma_{LT} = 0.3$, $\beta = 4$, $\rho = -0.5$, $r = 0.05$. Figure 5 displays the results of our numerical study. We call $S^*_{A,t}$ the early exercise boundary computed under assumption i), and $S^*_{E,t}$ the one under assumption ii). The two boundaries $S^*_{A,t}$ and $S^*_{E,t}$ coincide after the dividend at $t = 0.25$ is paid out, since there are no more intermediate cash flows before maturity, and the continuation value is in both cases the one of a European call with a time-to-maturity of 3 months. At $t = 0$, the two boundaries are different, with $S^*_{E,t} < S^*_{A,t}$. This means that an investor basing his exercise decision on the approximation $S^*_{E,t}$, may suboptimally decide to exercise the option at $t = 0$, if $S^*_{E,0} < S_0 < S^*_{A,0}$. He would then incur a loss given by $C(S_0 - d, 0) - (S_0 - K)$, where $C(S_0 - d, 0)$ is the correct price of an American option, computed at the stock value $S_0 - d$. The upper bound for the loss coming from the escrowed dividend approximation is reached exactly when $S_0 = S^*_{E,0}$. In this case, the following formulae give the maximum percentage loss at $t = 0$ under the three modelling environments considered and with the model parameters given above:

$$DL\%_{BS} = \frac{C_{BS}(S^*_{E,BS}) - (S^*_{E,BS} - K)}{C_{BS}(S^*_{E,BS})} = 0.63\%,$$
$$DL\%_{MRT} = \frac{C_{MRT}(S^*_{E,MRT}) - (S^*_{E,MRT} - K)}{C_{MRT}(S^*_{E,MRT})} = 0.48\%,$$
$$DL\%_{BTS} = \frac{C_{BTS}(S^*_{E,BTS}) - (S^*_{E,BTS} - K)}{C_{BTS}(S^*_{E,BTS})} = 1.14\%,$$

where we have suppressed for readability the dependence from time, and where the indexes BS, MRT, BTS mean that we have computed the relevant quantities under the Black Scholes, Merton and Bates models, respectively. The investors can lose up to 1% of the market value of their option if they exercise according to the wrong boundary.

We finally check in our database if there are situations in which the stock reaches the value $S^*_{E,t}$ but not $S^*_{A,t}$ on the days before the ex-dividend dates. We find that this

occurrence does happen, and not sporadically. For example, on May 10th, 2006, the Dupont stock closes at 45.71 dollars. The call option with $K = 30$ and $T = 0.45$ should not be exercised if the continuation value is computed correctly, but the option should be exercised if the continuation value is approximated with a European price. In this case, if an investor exercises his option wrongly, he will suffer a loss given by: $DL\%_{BS} = 0.07\%$, $DL\%_{MRT} = 2\%$, $DL\%_{BTS} = 0.06\%$.

Appendix G. Comparison with Finite Difference Methods

In Section 3.3 of the main paper we compare the speed and accuracy in pricing American option on dividend-paying stocks of the recursive projections and of the finite difference scheme (FD) solution of the partial differential equation associated with the pricing problem. Our main reference in developing the FD scheme is in't Hout and Foulon (2010). Compared to our recursive projections, finite differences are a mature topic of research, and refinements to specific applications are continuously developed. We are not aware of works that extend in't Hout and Foulon (2010) to introduce discrete dividends in the same way that Vellekoop and Nieuwenhuis (2006) improve binomial trees. Even though we believe that the picture we give of the relative performance of the two methods is fair, we are aware that optimisations in the FD toolbox may attenuate the advantage of the recursive projections. In this section, we first compare the convergence properties of our implementation of the FD scheme with the one of in't Hout and Foulon (2010), and then we try to provide theoretical insights on why we are convinced that recursive projections provide a computational advantage when pricing American options on dividend-paying stocks.

As a first check of the quality of our implementation of the FD scheme, we can

compare the convergence results of Figure (6) of the main text with the ones of Figure 5 of in't Hout and Foulon (2010). In *FD* schemes, and in the alternating direction implicit (ADI) variant (see below) that we implement, the number of time steps needed to obtain convergence is a faithful measure of the efficiency of the implementation. Each time step is composed of few simple matrix operations. We use specialised libraries for sparse matrices to achieve these operations, and it is almost impossible to obtain substantial speed gains in this part of the algorithm. In their implementation, in't Hout and Foulon (2010) obtain a $1bp$ error for a discretisation parameter between 500 and 1000. According to their different simulation exercises on European options, the time steps needed to achieve convergence are independent of the time to maturity. It is reasonable to think that this is the number of time steps needed between each exercise date in our example. They achieve this convergence with a slightly lower gird points in the stock dimension ($m_s = 200$ instead of 400), but higher for the volatility ($m_v = 100$ instead of 31), so that the overall number of grid points is comparable. Their slight gain in terms of grid points is obtained at the cost of using a non equispaced grid, which makes the implementation with dividends more complicated. Let us assume that by fine tuning our *FD* scheme implementation, or by devising a more efficient way of treating the dividends, we could achieve a convergence within 1 *bp* with $L_T = 512$, which is the lower bound to obtain a $1bp$ convergence in in't Hout and Foulon (2010) in the European case. According to Figure (6) of the main text, the computational time would still be of at least $10s$, an order of magnitude slower than the recursive projections.

We now provide some theoretical arguments in favour of using the recursive projections over finite difference schemes when pricing American options on dividend-paying stocks. We organize our discussion in 4 points. In the following, N gives the number of points in which the function is evaluated. In the case that the underlying asset follows a multidimensional process, N is the product of the number of grid points in each

dimension.

1. *FD* methods can achieve quadratic convergence in the time discretization parameter Δt, provided that we implement an alternating direction implicit (ADI) scheme. ADI schemes mean that for each time step, we need to perform two or more intermediate time steps. In the same way that using central differences instead of one sided differences provides a quadratic convergence to the differential operator in the stock dimension, these intermediate time steps assure a quadratic convergence in the time dimension. In the implementation of Section 3.3, we assumed $\rho = 0$ in the dynamics of Equation (21). Under this assumption, we need one intermediate time step. Had we chosen $\rho \neq 0$, we would have needed one or two additional time steps (see in't Hout and Foulon (2010), Section 2.3). Since the time needed by the FD scheme to price the option is linear in the number of time steps, this would have led to multiplying the computation time by a factor of 1.5 or 2. The computation time needed by the recursive projections does not depend on ρ being zero. This is a first reason why our comparison is not biased towards the recursive projections.

2. Let us assume for the moment that the process of the underlying is purely diffusive. The complexity, i.e., the number of computations needed, of the ADI is of the order $O(L_T N)$, where L_T is the number of time steps. The complexity of the recursive projections is $O(LN^2)$, where L is the number of recursive steps, which, we remind, are much less than L_T. The relative efficiency of the two methods depends on the ratio L/L_T. In Section 3.3, in the example of the call option written on a stock that distributes 3 dividends before maturity, $L = 3$ while L_T may reach 2^{10} to have a $1bp$ convergence (see Figure 5 in the main text). Figure 6 shows why in this example the two methods of recursive projection and finite differences require a very different number of steps in the time dimension. In a situation when L becomes large, e.g. an American put option, the differences would be less marked.

3. To compute the matrices needed to achieve a time step in the *ADI*, one needs to invert matrices of dimension $N \times N$. In the example of the call option written on a stock that distributes 3 dividends before maturity, the matrices are band-diagonal (sparse), and this means that the inversion has a computational cost of order $O(N)$. The matrices need to be computed only once, and can be applied to all time steps. Thus, the overall $O(N)$ complexity of the *FD* remains unaffected. If the underlying has a jump component, the pricing equation we need to solve is a partial integral differential equation. The integral term gives rise to a dense matrix that we need to invert, raising the complexity of the algorithm to $O(N^3)$. In specific cases, we can keep the complexity to $O(N \log(N))$. For instance, d'Halluin et al. (2005) achieve a $O(N \log(N))$ complexity, but in order to do so, they need to implement an iterative method to obtain the inverse of the matrix. This procedure is complex and application specific. First, we need to know the density of the jump component analytically to be able to apply FFT techniques. Second, we can only apply this technique to deterministic volatility models. It is not clear whether the same results could apply to a stochastic volatility model. Moreover, the implementation of d'Halluin et al. (2005) cannot reach quadratic convergence in the time dimension with a regular grid in t. This raises issues for a possible application to the discrete dividend case. In the empirical Section 4.2, we show that including jumps to a stochastic volatility models is as simple as multiplying two transition matrices, without adding numerical complexity to the problem.

4. Applying parallel computing to the recursive projections is straightforward. Equation (18) shows that, in the stochastic volatility case, we obtain every value v_{ip} by an entry-wide product of two matrices. In principle, we could compute all the v_{ip} at once in parallel. This is not feasible with the *ADI* scheme, since we obtain the entire value function at time t_l as a matrix time vector multiplication involving the entire value function at time t_{l+1}.

The four points above should convince the reader that the recursive projections compare well with *ADI* methods in most applications, are better in others, and in each case are much easier to implement. In the exercise of Section 3.3, we compare the recursive projections against the *ADI* in a pure diffusive framework, which is the most favorable case for the finite differences schemes. As we depart from the simple diffusion process, the advantages of recursive projections become even clearer, first of all in the easiness of implementation and generality. Our process has similar generality to Monte Carlo methods in terms of scope of application. Since the recursive and projection steps disentangle the reconstruction of the value function from the operator that drives the dynamics of the system, changing a payoff function, or adding intermediate cash flows, does not make the recursive projections more complicated, both in numerical complexity and in practical implementation. We need to code *ADI* methods for each pricing problem at hand, since we have to take care of both payoff and intermediate cash flows in an ad hoc way. In the recursive projections, adding additional components to the dynamic, as for instance adding a jump component to a stochastic volatility model, is as simple as sampling transition densities, and multiplying the respective matrices. Doing the same with an *ADI* scheme is not trivial, as would in most cases mean developing a new, problem specific, scheme.

References

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

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.

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

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.

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

Cosma, Antonio, Olivier Scaillet, and Rainer von Sachs, 2007, Multivariate wavelet-based shape preserving estimation for dependent observations, *Bernoulli* 13, 301–329.

Dechevsky, Lubomir, and Spiridon Penev, 1997, On shape-preserving probabilistic wavelet approximators, *Stochastic Analysis and Applications* 15, 187–215.

d'Halluin, Yann, Peter A Forsyth, and Kenneth R Vetzal, 2005, Robust numerical methods for contingent claims under jump diffusion processes, *IMA Journal of Numerical Analysis* 25, 87–112.

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

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

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.

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.

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

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.

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.

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
MMM	0.25	1.10	0.20	0.11	-0.13	0.40	0.16	-0.12	0.68	0.29	1.38	-0.44	0.36
AA	0.38	1.68	0.30	0.22	-0.14	0.50	0.28	-0.18	0.88	0.36	1.61	-0.33	0.34
AXP	0.34	2.19	0.25	0.14	-0.11	0.58	0.14	-0.06	0.78	0.37	1.34	-0.54	0.33
T	0.27	1.05	0.21	0.15	-0.09	0.37	0.16	-0.11	0.69	0.33	1.48	-0.26	0.29
BAC	0.32	1.58	0.24	0.18	-0.16	1.20	0.18	-0.14	0.98	0.36	1.60	-0.45	0.34
BA	0.31	1.54	0.24	0.15	-0.13	0.40	0.18	-0.12	0.80	0.33	1.56	-0.37	0.24
CAT	0.32	1.51	0.26	0.13	-0.10	0.54	0.15	-0.06	0.78	0.35	1.46	-0.37	0.30
CHV	0.24	1.00	0.20	0.12	-0.09	0.26	0.15	-0.10	0.55	0.27	1.55	-0.26	0.23
CSCO	0.32	1.36	0.25	0.17	-0.12	1.53	0.08	-0.10	1.08	0.32	1.88	-0.36	0.29
KO	0.24	1.03	0.19	0.13	-0.12	0.41	0.15	-0.11	0.66	0.27	1.49	-0.34	0.23
XOM	0.24	0.92	0.19	0.15	-0.12	0.79	0.17	-0.12	0.69	0.24	1.57	-0.38	0.27
GE	0.27	1.06	0.21	0.17	-0.14	0.41	0.18	-0.17	0.83	0.36	1.43	-0.30	0.34
HWP	0.37	1.81	0.28	0.20	-0.13	0.64	0.24	-0.20	0.98	0.45	1.76	-0.33	0.31
HD	0.32	1.45	0.24	0.21	-0.15	0.43	0.28	-0.19	0.77	0.38	1.77	-0.39	0.34
INTC	0.38	1.82	0.29	0.20	-0.14	0.38	0.30	-0.28	0.75	0.36	1.64	-0.32	0.26
IBM	0.28	1.81	0.21	0.13	-0.14	0.51	0.22	-0.15	0.71	0.29	1.84	-0.38	0.21
JNJ	0.22	0.87	0.17	0.13	-0.10	0.32	0.16	-0.10	0.67	0.25	1.49	-0.29	0.24
JPM	0.33	1.19	0.28	0.15	-0.10	0.27	0.18	-0.06	0.64	0.33	1.61	-0.34	0.26
MCD	0.25	1.08	0.20	0.13	-0.11	0.32	0.12	-0.12	0.65	0.26	1.31	-0.37	0.25
MRK	0.27	1.22	0.22	0.15	-0.11	0.40	0.14	-0.12	0.80	0.36	1.58	-0.32	0.24
MSFT	0.25	1.34	0.19	0.17	-0.09	0.35	0.24	-0.13	0.77	0.28	1.56	-0.22	0.34
PFE	0.28	1.45	0.21	0.17	-0.10	0.44	0.20	-0.13	0.80	0.28	1.25	-0.20	0.27
PG	0.21	1.00	0.17	0.14	-0.11	0.74	0.11	-0.07	0.57	0.25	1.29	-0.39	0.22
TRV	0.29	1.41	0.21	0.17	-0.08	0.49	0.16	-0.05	0.82	0.32	1.71	-0.23	0.27
UNH	0.33	1.30	0.27	0.18	-0.16	0.99	0.24	-0.17	0.93	0.32	1.55	-0.48	0.28
UTX	0.27	1.17	0.22	0.13	-0.11	0.38	0.16	-0.11	0.66	0.30	1.47	-0.38	0.26
VZ	0.28	1.24	0.21	0.18	-0.11	0.60	0.17	-0.09	0.74	0.33	1.41	-0.24	0.33
WMT	0.26	1.14	0.21	0.15	-0.09	0.40	0.19	-0.09	0.71	0.29	1.44	-0.31	0.26
DIS	0.29	1.23	0.22	0.16	-0.09	0.50	0.23	-0.05	0.74	0.31	1.44	-0.37	0.26
DD	0.28	1.25	0.22	0.15	-0.12	0.37	0.18	-0.13	0.67	0.28	1.51	-0.40	0.27

Table 1: 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 average values shown in the table are computed on the results of those 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.

*The source of the calibrated parameters of the SP500 dynamics is the work of Bakshi, Cao and Chen (1997).

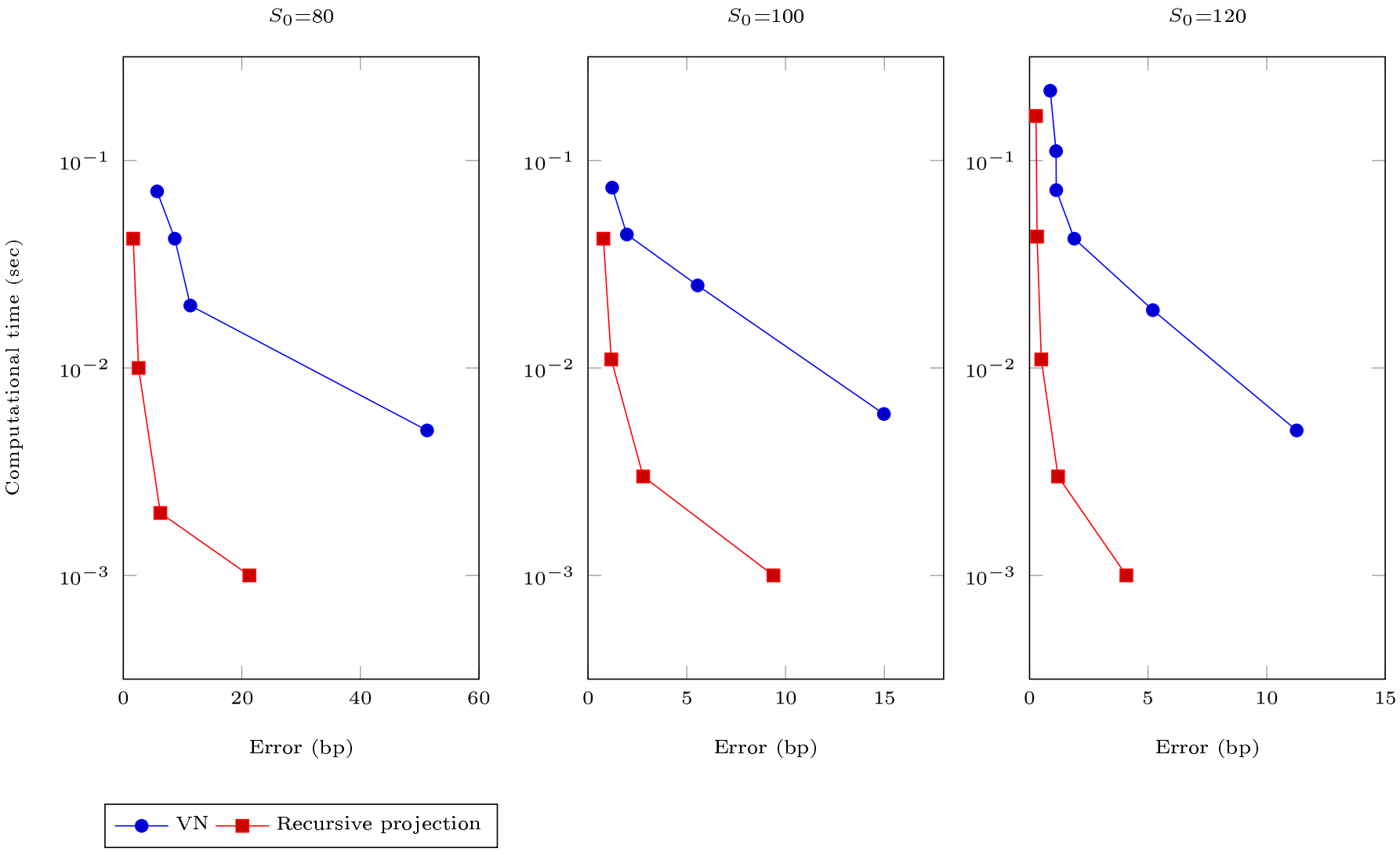

Fig. 1. Comparison between the approximated binomial tree method (VN) of Vellekoop and Nieuwenhuis (2009) and the recursive projections on an American call option written on a dividend-paying stock in the Black-Scholes case. The option has a maturity of 3 years and a dividend $d = 2$ is paid at the end of each year. Other parameters set equal to $r = 0.05, \sigma = 0.2, K = 100$.

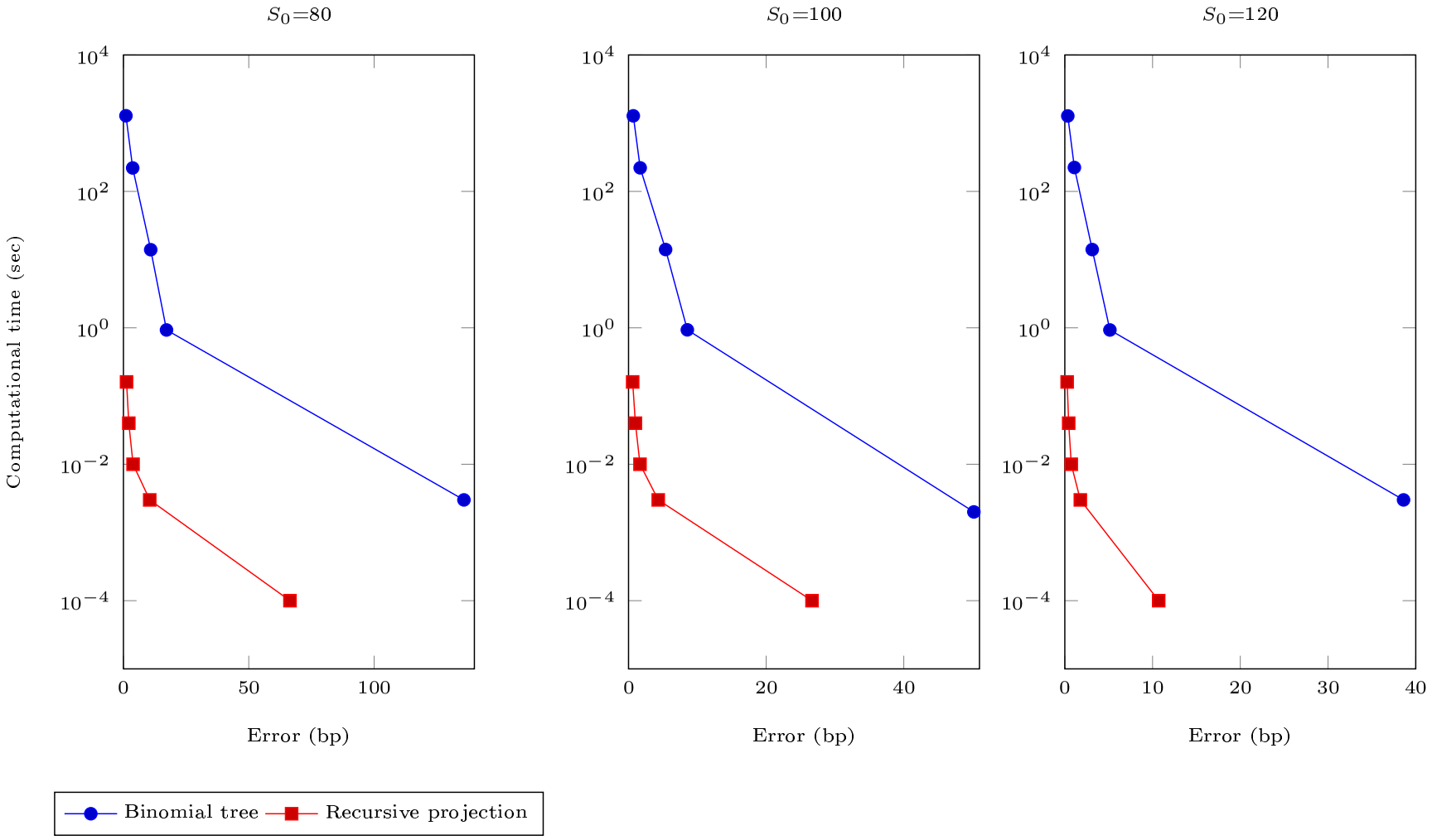

Fig. 2. Comparison between the binomial tree and the recursive projection method on an American call option written on a dividend-paying stock in the Black-Scholes case. The option has a maturity of 3 years and a dividend $d = 2$ is paid at the end of each year. Other parameters are set equal to $r =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

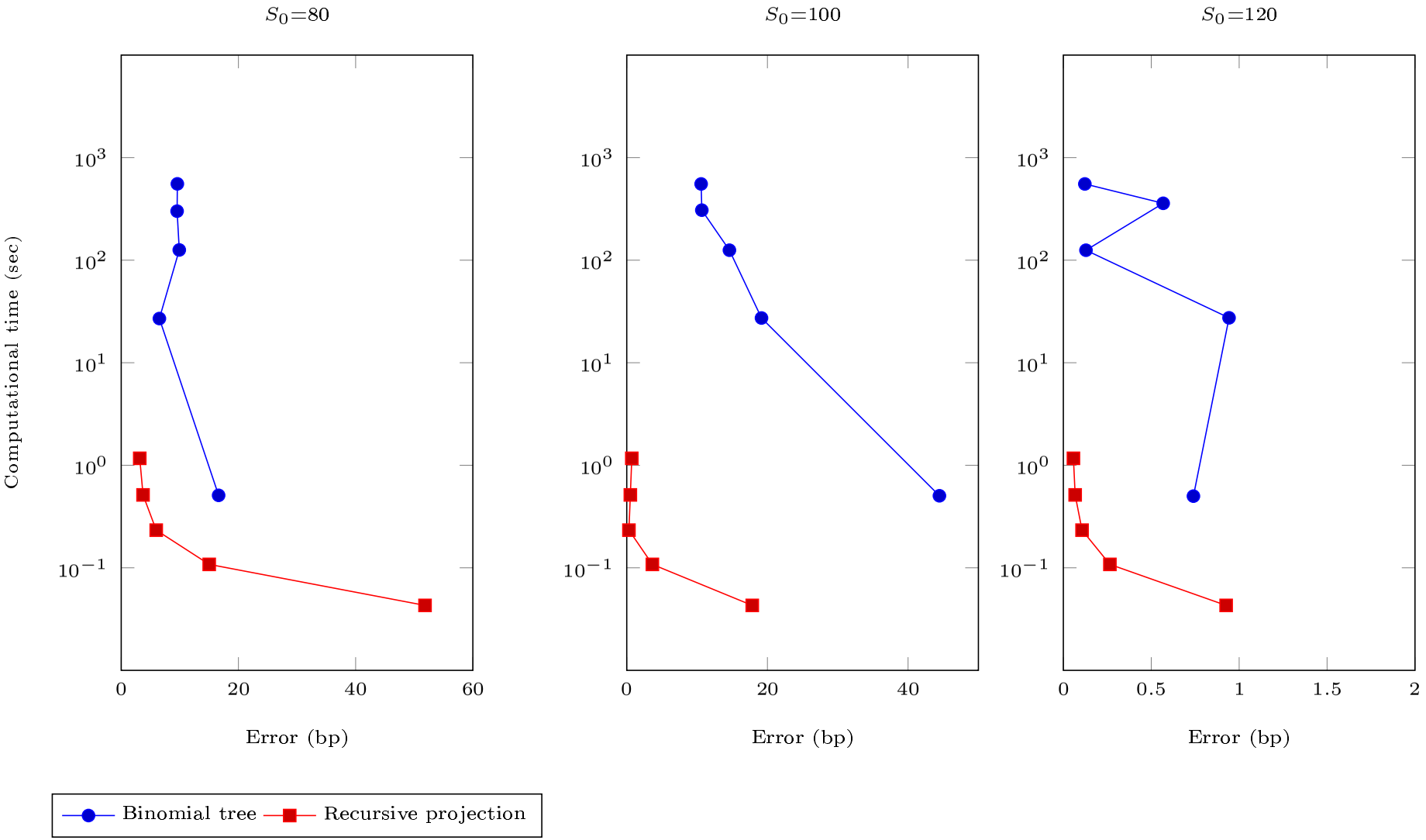

Fig. 3. Comparison between the binomial tree and the recursive projection method on a Bermudan digital call option in the Black-Scholes case. The option has a maturity of 10 years and can be exercised 4 times per year. Other parameters are set equal to $r = 0.1$, $\sigma = 0.2$, and $K = 100$.

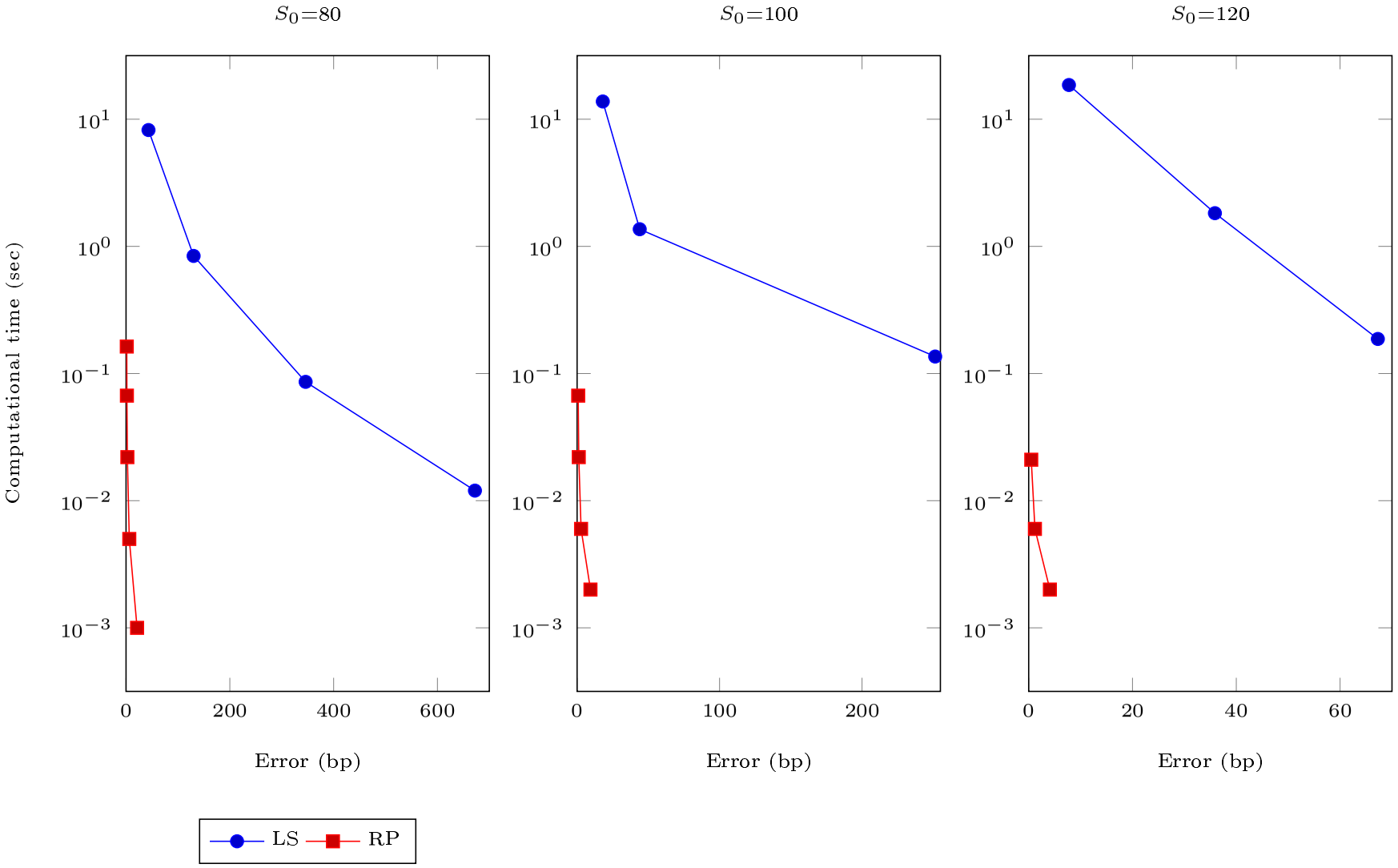

Fig. 4. Comparison between the simulation least-square method (LS) of Longstaff and Schwartz (2001) and our recursive projection method (RP) on an American call option written on a dividend paying stock in the Black-Scholes case. The option has a maturity of 3 years and a dividend $d = 2$ is paid at the end of each year. Other parameters set equal to $r = 0.05, \sigma = 0.2, K = 100$.

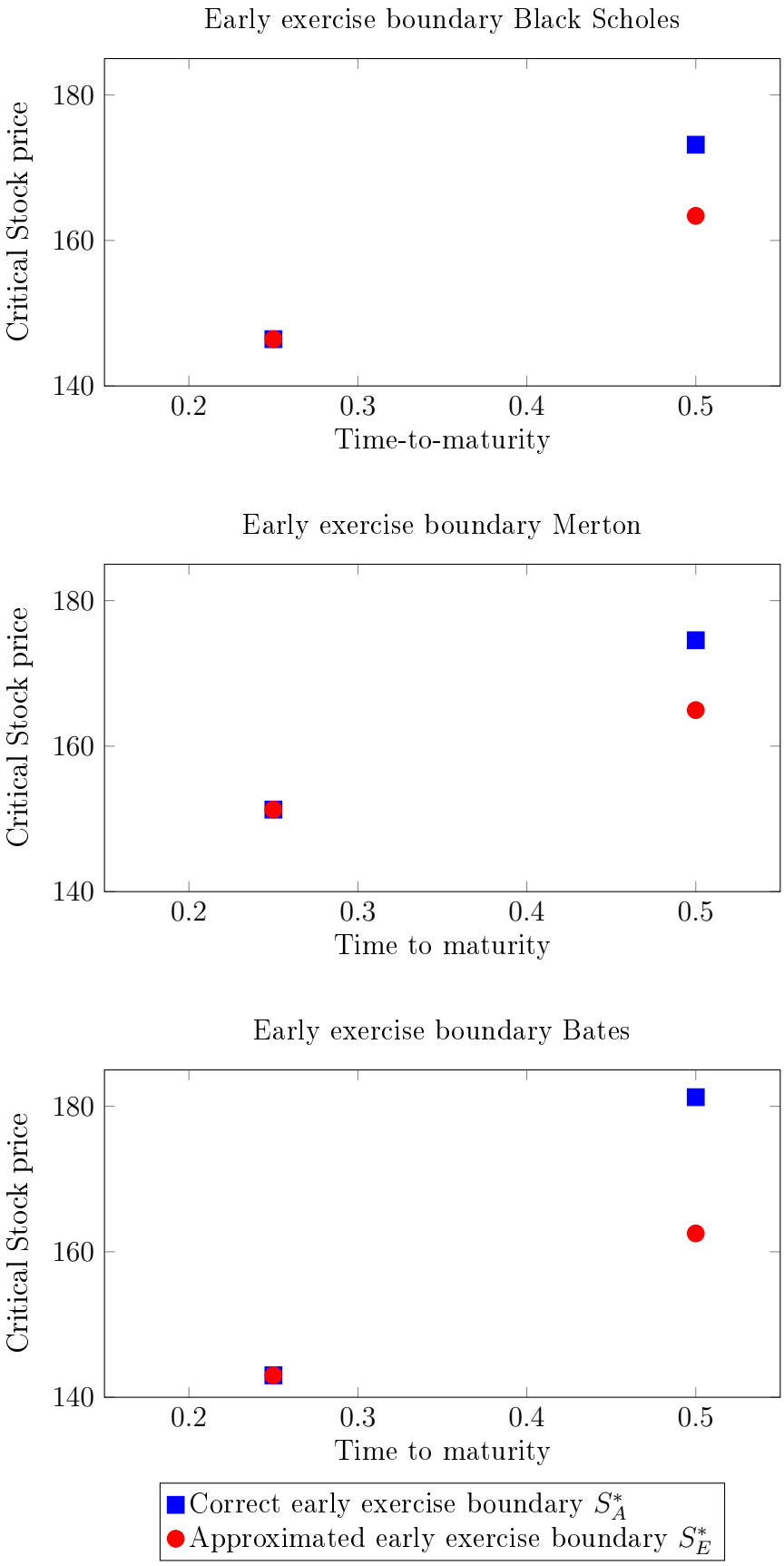

Fig. 5. Comparison between the true early exercise boundary S_A^* and the approximated early exercise boundary S_E^*. S_E^* is calculated by approximating the continuation value of the option with the price of a European option where the starting value of the stock is set equal to S_0 minus the present value of all future dividends.

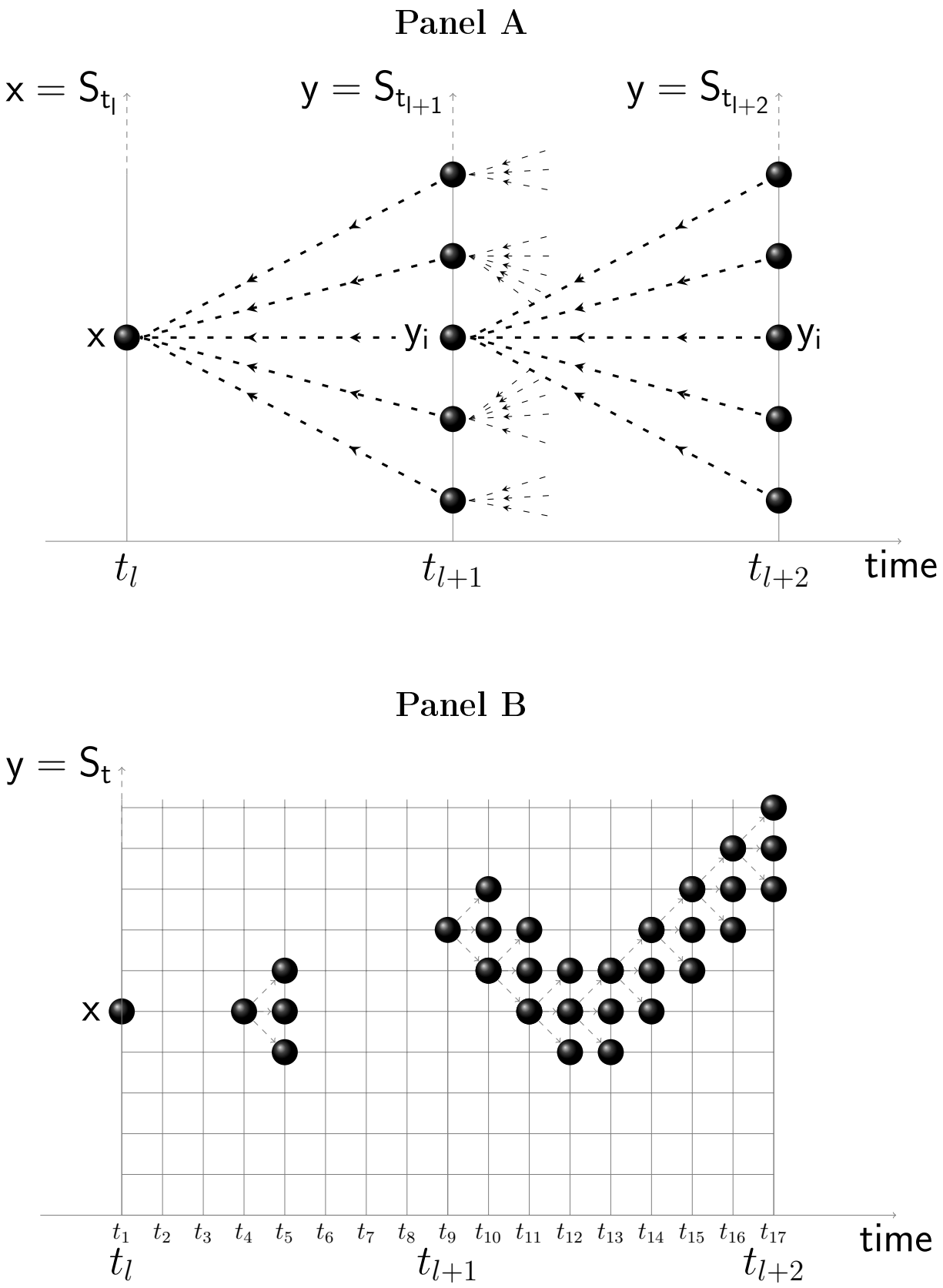

Fig. 6. Comparison between recursive projections and *FD* scheme. In both methods the value function $V(S_t, t)$ is computed on a time-homogeneous equally spaced grid. Recursive projections need only to compute $V(S_t, t)$ at specific times t_l, t_{t+1}, t_{l+2} (Panel A), while *FD* need to compute $V(S_t, t)$ on a finer grid to achieve convergence (Panel B).