The Annals of Applied Probability
2011, Vol. 21, No. 2, 589–608
DOI: 10.1214/10-AAP704
© Institute of Mathematical Statistics, 2011

THE LONGSTAFF–SCHWARTZ ALGORITHM FOR LÉVY MODELS: RESULTS ON FAST AND SLOW CONVERGENCE

BY STEFAN GERHOLD[1]

Vienna University of Technology

We investigate the Longstaff–Schwartz algorithm for American option pricing assuming that both the number of regressors and the number of Monte Carlo paths tend to infinity. Our main results concern extensions, respectively, applications of results by Glasserman and Yu [*Ann. Appl. Probab.* **14** (2004) 2090–2119] and Stentoft [*Manag. Sci.* **50** (2004) 1193–1203] to several Lévy models, in particular the geometric Meixner model. A convenient setting to analyze this convergence problem is provided by the Lévy–Sheffer systems introduced by Schoutens and Teugels.

1. Introduction. PDE or tree methods for pricing financial products become ineffective in the presence of many stochastic factors and path dependent payoff structures. When resorting to Monte Carlo, early exercise features like callability or flip options pose difficulties. Typical examples are the pricing of callable LIBOR exotics with the LIBOR market model [3] or the valuation of life insurance contracts with early exercise features [1].

The least squares Monte Carlo approach by Longstaff and Schwartz [17] has become the standard method to deal with such American/Bermudan products. It proceeds by backward induction and estimates value functions by regression on a prescribed set of basis functions. The computed exercise strategy is suboptimal, resulting in a lower bound for the option price; see Belomestny, Bender and Schoenmakers [2] for recent work on upper bounds. Fouque and Han [12] discuss numerical aspects of American option pricing, including variance reduction.

The convergence analysis of the Longstaff–Schwartz algorithm was commenced in the original paper [17] and was carried out in detail by Clément, Lamberton and Protter [4]. They show convergence of the regression approximation to the true Bermudan price and convergence of the Monte Carlo procedure for a fixed number of basis functions. Glasserman and Yu [14] and Stentoft [25] have analyzed settings in which the number of basis functions and the number of simulation paths increase together. In particular, Glasserman and Yu [14] have shown

Received March 2008; revised March 2010.

[1]Supported by the Christian Doppler Laboratory for Portfolio Risk Management (PRisMa Lab) http://www.prismalab.at/. The Bank Austria Creditanstalt (BA-CA) and the Austrian Federal Financing Agency (ÖBFA) through CDG is gratefully acknowledged.

MSC2010 subject classifications. Primary 62P05; secondary 33C45.

Key words and phrases. Option pricing, dynamic programming, Monte Carlo, regression, orthogonal polynomials, Lévy–Meixner systems.

that the number of paths must grow exponentially in the number of basis functions if the underlying process is Brownian motion or geometric Brownian motion. On the other hand, Stentoft [25] appealed to results on series estimators [7, 20] to obtain polynomial growth for rather general models, assuming that the underlying has a bounded state space. The latter assumption was also imposed by Eglof, Kohler and Todorovic [9, 10] in the analysis of their extension of the Longstaff–Schwartz algorithm.

In the present paper we discuss the applicability of Stentoft's results to exponential Lévy models and extend Glasserman and Yu's analysis to several models, including the Meixner model [16, 22]. These latter results provide an application of the neat martingale properties that Schoutens and Teugels [21, 23] found for certain Lévy processes and families of orthogonal polynomials.

In the following section we recall the dynamic programming principle and the Longstaff–Schwartz algorithm. We show how Stentoft's [25] convergence result can be applied to Lévy models, in particular, to the Meixner model. This involves discussing the assumption of a bounded underlying and the smoothness of the value functions occurring in the backward induction.

In Section 3 we describe the problem that Glasserman and Yu [14] treated. The main difference to Stentoft's setting is the unbounded support of the underlying. Section 4 recalls the notions of Sheffer system and Lévy–Meixner system. Besides Brownian motion, this theory yields four processes that lend themselves to the investigation: the Meixner, standard Poisson, Gamma and Pascal processes [8, 13]. In Section 5 we assume that our option has only three exercise opportunities resulting in a single regression and show how fast the number of simulation paths must increase in order to ensure convergence of the Longstaff–Schwartz algorithm for a growing number of basis functions. Finally, Section 6 contains an analogous bound for the multi-period setting, which is weaker, but upon inversion still leads to the same critical asymptotic rate as the single-period case. In the course of the proofs it turns out that the different critical rate pertaining to Brownian motion stems from the comparatively slow growth of the linearization coefficients of the associated Lévy–Meixner system, namely, the Hermite polynomials.

2. Bounded state space and fast convergence. Suppose that our asset follows a Markov process S_t. We assume throughout the paper that the interest rate is zero; extending our results to a constant interest rate $r > 0$ is trivial. Consider a Bermudan option (which may serve as a proxy for an American option) that can be exercised at the times $0 = t_0 < \cdots < t_m$. The payoff from exercise is $h_n(S_{t_n})$ for given functions h_n, $0 \le n \le m$. By the dynamic programming principle the option value at time $t_0 = 0$ equals $V_0 = \max\{h_0(S_0), C_0(S_0)\}$, where the continuation values C_n are given by

$$C_m(x) = 0,$$

$$C_n(x) = \mathbf{E}[\max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}(S_{t_{n+1}})\} \mid S_{t_n} = x], \qquad 0 \le n < m.$$

Suppose that N sample paths of the underlying are simulated. Longstaff and Schwartz [17] propose to approximate the continuation values by a linear combination of basis functions ψ_{nk},

$$C_n(x) \approx \sum_{k=0}^{K} \beta_{nk}\psi_{nk}(x) = \beta_n^{\mathrm{T}}\psi_n(x),$$

where $\beta_n = (\beta_{n0}, \ldots, \beta_{nK})^{\mathrm{T}}$ is a vector of real numbers which is estimated by regression over the simulated paths and $\psi_n(x) = [\psi_{n0}(x), \ldots, \psi_{nK}(x)]^{\mathrm{T}}$.

To obtain a good convergence result as N and K both tend to infinity, Stentoft [25] assumes that samples above and below certain thresholds are discarded. So let us fix finite truncation intervals $I_1, \ldots, I_m \subset]0, \infty[$ and discard all sample paths with $S_{t_n} \notin I_n$ when estimating the continuation value $C_n(x)$. We are then estimating the following "truncated" continuation values:

$$\begin{aligned}
&C_m^{\mathrm{tr}}(x) = 0,\\
&C_n^{\mathrm{tr}}(x) = \mathbf{1}_{I_n}(x) \cdot \mathbf{E}[\max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}^{\mathrm{tr}}(S_{t_{n+1}})\} \mid S_{t_n} = x], \qquad 1 \le n < m,\\
&C_0^{\mathrm{tr}}(x) = \mathbf{E}[\max\{h_1(S_{t_1}), C_1^{\mathrm{tr}}(S_{t_1})\} \mid S_0 = x].
\end{aligned}$$

The option value at time $t_0 = 0$ approximately equals

$$V_0^{\mathrm{tr}} = \max\{h_0(S_0), C_0^{\mathrm{tr}}(S_0)\}. \tag{2.1}$$

Outside of the truncation intervals $I_n \subset]0, \infty[$ we extrapolate by zero since it does not matter in the theoretical analysis.

Besides truncation, another possibility to make the state space bounded would be absorption of the underlying process at some lower and upper bounds [9, 10]. This, however, causes atoms in the distribution so that Stentoft's result is no longer applicable as it requires the existence of a density.

We assume in the present section that the underlying has the following dynamics. (Recall that we suppose throughout that the interest rate is zero.)

ASSUMPTION A (Exponential Lévy dynamics). The risk neutral dynamics of the underlying are

$$S_t = S_0 \exp(X_t),$$

where X_t is a Lévy process with $X_0 = 0$. The support of X_t is the whole real line for $t > 0$ and X_t has a continuous density function.

ASSUMPTION B (Value smoothness). Let the function h be of, at most, linear growth and such that $h(S_T)$ is integrable for each $T > 0$. Then $\mathbf{E}[h(S_T) \mid S_0 = x]$ is a C^1-smooth function of x.

Without going into detail we note that Stentoft [25] imposes the following additional assumptions:

ASSUMPTION C (Further technical assumptions). The basis functions are shifted Legendre polynomials, the continuation values $C_n(S_{t_n})$ are in the L^2-span of the regressors, the simulated paths are independent and the probability that the exercise payoff exactly equals the continuation value is zero.

Now Stentoft's main result ([25], Theorem 2), specialized to Lévy models, reads as follows. (By "truncated algorithm" we mean that we discard the samples outside the intervals I_n as explained above.)

THEOREM 1. *Fix arbitrary finite truncation intervals $I_1, \dots, I_m$ contained in $]0, \infty[$ and assume that Assumptions A–C hold. Let N (the number of paths) and K (the number of basis functions) tend to infinity such that $K^3/N \to 0$. Then the option prices computed by the truncated Longstaff–Schwartz algorithm converge to V_0^{tr}, defined by (2.1).*

If the truncation intervals are large enough, then one would hope that the approximate price V_0^{tr} is close to the exact price V_0. We will now show that this is indeed the case for Lévy models, assuming mild integrability and (at most) linearly growing payoff functions.

ASSUMPTION D (Integrability). For each t there are $p > 1$ and $p' > 0$ such that S_t^p and $S_t^{-p'}$ are integrable.

ASSUMPTION E (Linear payoff growth). The payoff functions grow at most linearly,

$$|h_n(x)| \le c(1+x), \qquad x \ge 0, 1 \le n \le m, \text{ for some } c > 0. \tag{2.2}$$

THEOREM 2. *Assume that Assumptions A, D and E are satisfied and that the truncation intervals satisfy*

$$I_n = [b_n^{-1}, b_n], \qquad 1 \le n < m,$$

where

$$b_n = b_{n+1}^{\nu}, \qquad 1 \le n < m-1, \tag{2.3}$$

with

$$\nu = \min\left\{\frac{p'}{p'+q}, \frac{p}{p+q}\right\} \quad \textit{and} \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Then V_0^{tr} converges to the exact option price V_0 as b_m tends to infinity.

Note that $\nu = 1 - 1/p$ in (2.3) if $p = p'$. In particular, if all moments of the underlying and its reciprocal exist, like in the Black–Scholes model, then the exponent $\nu \in]0, 1[$ is arbitrarily close to 1.

PROOF OF THEOREM 2. A trivial induction, using the martingale property of S_t, shows that the continuation values $C_n(x)$ and $C_n^{\mathrm{tr}}(x)$ satisfy the bound (2.2) too. We will show that for all n

$$C_n^{\mathrm{tr}}(x) = C_n(x) + o(1) \qquad \text{as } b_m \to \infty, \text{ uniformly w.r.t. } x \in I_n.$$

For $x \in I_n$, we have

$$\begin{aligned}
(2.4)\quad & C_n(x) - C_n^{\mathrm{tr}}(x) \\
&= \mathbf{E}\big[\mathbf{1}_{\{S_{t_{n+1}} \notin I_{n+1}\}} \max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}(S_{t_{n+1}})\} \mid S_{t_n} = x\big] \\
&\quad + \mathbf{E}\big[\mathbf{1}_{\{S_{t_{n+1}} \in I_{n+1}\}} \max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}(S_{t_{n+1}})\} \mid S_{t_n} = x\big] \\
(2.5)\quad &\quad - \mathbf{E}\big[\mathbf{1}_{\{S_{t_{n+1}} \in I_{n+1}\}} \max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}^{\mathrm{tr}}(S_{t_{n+1}})\} \mid S_{t_n} = x\big] \\
&\quad - \mathbf{E}\big[\mathbf{1}_{\{S_{t_{n+1}} \notin I_{n+1}\}} \max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}^{\mathrm{tr}}(S_{t_{n+1}})\} \mid S_{t_n} = x\big].
\end{aligned}$$

It follows readily from the induction hypothesis that the difference of the second and the third term is uniformly $o(1)$ on I_n, as $b_{m-1} \to \infty$. In the following, we write c for various positive constants whose precise value is irrelevant. Now let us estimate the first and the last expectation on the right-hand side of (2.5). Again, for $x \in I_n$ we use Hölder's inequality and Minkowski's inequality to see that each of them is bounded by

$$\begin{aligned}
& \mathbf{E}\big[\mathbf{1}_{\{S_{t_{n+1}} \notin I_{n+1}\}} c(1 + S_{t_{n+1}}) \mid S_{t_n} = x\big] \\
&\quad \le c\mathbf{P}[S_{t_{n+1}} \notin I_{n+1} \mid S_{t_n} = x]^{1/q} \cdot \mathbf{E}[(1 + S_{t_{n+1}})^p \mid S_{t_n} = x]^{1/p} \\
(2.6)\quad &\quad = c\mathbf{P}\left[x \frac{S_{t_{n+1}}}{S_{t_n}} \notin I_{n+1}\right]^{1/q} \cdot \mathbf{E}\left[\left(1 + x\frac{S_{t_{n+1}}}{S_{t_n}}\right)^p\right]^{1/p} \\
&\quad \le c\left(1 - F_n\left(\frac{b_{n+1}}{x}\right) + F_n\left(\frac{1}{x b_{n+1}}\right)\right)^{1/q} \left(1 + x\mathbf{E}\left[\left(\frac{S_{t_{n+1}}}{S_{t_n}}\right)^p\right]^{1/p}\right) \\
&\quad \le c b_n \left(1 - F_n\left(\frac{b_{n+1}}{b_n}\right) + F_n\left(\frac{b_n}{b_{n+1}}\right)\right)^{1/q},
\end{aligned}$$

where F_n is the distribution function of $S_{t_{n+1}}/S_{t_n}$. Now note that

$$\begin{aligned}
b_n^q\left(1 - F_n\left(\frac{b_{n+1}}{b_n}\right)\right) &= b_n^{p+q} b_{n+1}^{-p} \left(\frac{b_{n+1}}{b_n}\right)^p \left(1 - F_n\left(\frac{b_{n+1}}{b_n}\right)\right) \\
&\le \left(\frac{b_{n+1}}{b_n}\right)^p \left(1 - F_n\left(\frac{b_{n+1}}{b_n}\right)\right) \\
&= o(1), \qquad b_{m-1} \to \infty,
\end{aligned}$$

where the last equality follows [11] from $S_{t_{n+1}}/S_{t_n} \in L^p$. Similarly, if G_n denotes the distribution function of $S_{t_n}/S_{t_{n+1}}$, we have

$$\begin{aligned} b_n^q F_n\left(\frac{b_n}{b_{n+1}}\right) &= b_n^q\left(1 - G_n\left(\frac{b_{n+1}}{b_n}\right)\right) \\ &= b_n^{p'+q} b_{n+1}^{-p'} \left(\frac{b_{n+1}}{b_n}\right)^{p'} \left(1 - G_n\left(\frac{b_{n+1}}{b_n}\right)\right) \\ &= o(1). \end{aligned}$$ □

Besides the bounded state space, a crucial assumption of Stentoft's result (Theorem 1) is the smoothness of the continuation value functions. In the Black–Scholes model, and more generally in models where the log-price X_t has a diffusion component, they are always C^∞-smooth [5]. The variance Gamma model is an example of a pure jump process where the value functions are not necessarily continuously differentiable [5]. In the geometric Meixner model [16, 22, 23], on the other hand, the continuation values are smooth, as we will now show. Consequently, Theorem 1 is applicable to the geometric Meixner model (if the mild Assumptions C and D are satisfied).

PROPOSITION 3. *Suppose that Assumptions A and D hold and that the log-price X_t is a Meixner process. Then Assumption B holds.*

PROOF. For fixed $t > 0$ the log-price X_t follows the Meixner distribution $\mathrm{Meix}(\alpha, \beta, \mu t, \delta t)$, where $\alpha > 0$, $-\pi < \beta < \pi$, $\mu > 0$ and $\delta \in \mathbb{R}$. This means that the density of X_t equals

$$f_t(x) = \frac{(2\cos(\beta/2))^{2\delta t}}{2\pi\alpha\Gamma(2\delta t)} e^{\beta/\alpha(x-\mu t)} \left|\Gamma\left(\delta t + i\frac{x-\mu t}{\alpha}\right)\right|^2$$

and the value function for the payoff $h(S_T)$ is

$$\begin{aligned} \mathbf{E}[h(S_T) \mid S_t = x] &= \int_{-\infty}^{\infty} h(e^y x) f_{T-t}(y)\, dy \\ &= \int_0^\infty h(z) f_{T-t}\left(\log\frac{z}{x}\right) dz/z. \end{aligned} \tag{2.7}$$

By the asymptotic formulas [22]

$$f_t(x) \sim c_\pm |x|^{2\delta t - 1} e^{-|x|(\pi \pm \beta)/\alpha}, \qquad x \to \pm\infty,$$

and the integrability Assumption D, we must have $(\pi + \beta)/\alpha > 1$. We can now differentiate the value function (2.7) under the integral sign, justified by the following fact: for real u and natural k the quantity

$$\frac{\partial^k/\partial v^k|\Gamma(u+iv)|}{|\Gamma(u+iv)|}$$

grows only polynomially in v as $v \to \pm\infty$. To see this start from Lerch's formula [15]

$$|\Gamma(u+iv)| = \frac{\Gamma(u+1)}{\sqrt{u^2+v^2}} \prod_{n=1}^{\infty} \left(1 + \frac{v^2}{(u+n)^2}\right)^{-1/2},$$

hence, we have

$$\frac{\partial/\partial v|\Gamma(u+iv)|}{|\Gamma(u+iv)|} = -\frac{v}{u^2+v^2} - v\sum_{n=1}^{\infty} \frac{1}{(u+n)^2+v^2}.$$

It suffices to note that $1/[(u+n)^2+v^2] \le 1/(u+n)^2$ to see that this expression grows only polynomially in v. The higher derivatives can be dealt with by a straightforward induction. $\square$

3. Unbounded state space and slow convergence. If we drop the assumption that the state space of our underlying is bounded, the convergence behavior of the Longstaff–Schwartz algorithm radically changes. (As above, we suppose that both the number of paths and the number of basis functions tend to infinity.) This is illustrated by results of Glasserman and Yu [14] who showed, assuming that the underlying follows either Brownian motion or geometric Brownian motion, that the number of Monte Carlo paths must grow exponentially in the number of basis functions to retain convergence. The first and last lines of Table 1 reflect this result; the lines in between will be established below.

For the reader's convenience, our notation closely follows that of [14]. Recall that we assume that the interest rate is $r = 0$ throughout the paper.

The variant of the Longstaff–Schwartz algorithm to be analyzed proceeds as follows. Start with the final continuation value $\hat{C}_m = 0$ and the final option value

TABLE 1

The highest possible number of basis functions for N paths

Process	**Basis polynomials**	**#Basis functions**
Geometric Brownian motion	Monomials	$\sqrt{\log N}$
Meixner	Meixner–Pollaczek	$\log N/\log\log N$
Standard Poisson	Charlier	$\log N/\log\log N$
Gamma	Laguerre	$\log N/\log\log N$
Pascal	Meixner	$\log N/\log\log N$
Brownian motion	Hermite	$\log N$

$\hat{V}_m = h_m$. For $n = m-1, \dots, 1$ generate N sample paths $\{S_{t_1}^{(i)}, \dots, S_{t_{n+1}}^{(i)}\}$, $1 \le i \le N$, and set

$$\begin{aligned}
\hat{\gamma}_n &= \frac{1}{N}\sum_{i=1}^{N} \hat{V}_{n+1}(S_{t_{n+1}}^{(i)})\psi_n(S_{t_n}^{(i)}),\\
\hat{\beta}_n &= \Psi_n^{-1}\hat{\gamma}_n,\\
\hat{C}_n &= \hat{\beta}_n^{\mathrm{T}}\psi_n,\\
\hat{V}_n &= \max\{h_n, \hat{C}_n\}.
\end{aligned}$$

Finally, the initial continuation value is $\hat{C}_0(S_0) = N^{-1}\sum_{i=1}^{N} \hat{V}_1(S_{t_1}^{(i)})$ from which the initial option value is estimated by $\hat{V}_0 = \max\{h_0(S_0), \hat{C}_0(S_0)\}$.

There are two (minor) differences to the variant of the algorithm that we analyzed in Section 2: first, we assume now that a fresh set of paths is generated for each exercise date. Second, in the present section we will use explicit expressions for the $(K+1)\times(K+1)$ matrix

$$\Psi_n = \mathbf{E}[\psi_n(S_{t_n})\psi_n(S_{t_n})^{\mathrm{T}}], \tag{3.1}$$

which has to be estimated by its sample counterpart in general.

In the single-period case $m = 2$, the question that Glasserman and Yu [14] treated is as follows. Suppose that there is an *exact* representation

$$h_2(S_{t_2}) = \sum_{k=0}^{K} \beta_k \psi_{2k}(S_{t_2}), \tag{3.2}$$

with unknown constants β_k. This assumption is not too restrictive; an infinite series representation of this kind has to be assumed anyway to get convergence of the algorithm and since we are interested in $K \to \infty$, we can suppose that (3.2) is a good approximation of the payoff at t_2. Furthermore, assume that the martingale property

$$\mathbf{E}[\psi_{2k}(S_{t_2}) \mid S_{t_1}] = \psi_{1k}(S_{t_1}) \tag{3.3}$$

holds. (In [14], additional deterministic factors in (3.3) are allowed; we chose to absorb these into the basis functions.) How fast may K tend to infinity compared to N while assuring that the mean square error of β tends to zero? To this end, Glasserman and Yu [14] established the bounds

$$\sup_{|\beta|=1} \mathbf{E}[|\beta - \hat{\beta}|^2] \le \frac{\|\Psi_1^{-1}\|^2}{N}\sum_{j=0}^{K}\sum_{k=0}^{K}\mathbf{E}[\psi_{2j}(S_{t_2})^2\psi_{1k}(S_{t_1})^2] \tag{3.4}$$

and

$$\sup_{|\beta|=1} \mathbf{E}[|\beta - \hat{\beta}|^2] \ge \frac{1}{N\|\Psi_1\|^2}\sum_{k=0}^{K}\mathbf{E}[\psi_{2K}(S_{t_2})^2\psi_{1k}(S_{t_1})^2] - \frac{1}{N}. \tag{3.5}$$

Here and in what follows, $|\cdot|$ denotes the Euclidean vector norm and $\|\cdot\|$ denotes the Euclidean (or Frobenius) matrix norm. With regard to notation, Glasserman and Yu [14] call the coefficients in (3.2) a_k instead of β_k; our simplified assumption (3.3) makes both their a and β equal to our β. This has to be kept in mind when comparing (3.4) and (3.5) to [14], formulas (22), respectively, (23).

The proofs of the estimates (3.4) and (3.5) are short; the bulk of the work of Glasserman and Yu [14] lies in the concrete examples (Brownian motion and geometric Brownian motion) and in the general analysis of the multi-period case on which we will build in Section 6.

The martingale property (3.3) is convenient for estimating the expectations in the bounds (3.4) and (3.5). Another useful property is orthogonality of the basis functions. If S_t is Brownian motion, then Glasserman and Yu [14] have shown that for N paths the highest K, for which the mean square error tends to zero, is roughly $\log N$. Hermite polynomials are natural basis functions in this case. If the underlying process is geometric Brownian motion and monomials are used as basis functions, then K may only be as high as $\sqrt{\log N}$. In the following sections we show that the analogous rate for the Meixner, Poisson, Gamma and Pascal processes is in between, namely, $\log N/\log\log N$.

4. Lévy–Meixner systems. A source of basis functions and processes that satisfy martingale equalities of the type (3.3) are the Lévy–Meixner systems introduced by Schoutens and Teugels [21, 23]. Recall that Meixner [18] has determined all sets of orthogonal polynomials $Q_k(x)$ that satisfy Sheffer's condition

$$f(z)\exp(xu(z)) = \sum_{k=0}^{\infty} Q_k(x)\frac{z^k}{k!}$$

for some formal power series f and u with $u(0) = 0$, $u'(0) \neq 0$ and $f(0) \neq 0$. Schoutens and Teugels [23] introduce a time parameter t via

$$f(z)^t \exp(xu(z)) = \sum_{k=0}^{\infty} Q_k(x,t)\frac{z^k}{k!}$$

and show how an infinitely divisible characteristic function, and thus a Lévy process, can be defined by f and u under appropriate conditions. Building on Meixner's characterization, five sets of orthogonal polynomials $Q_k(X_t,t)$ and associated Lévy processes X_t are determined which satisfy martingale equalities of the type

$$\mathbf{E}[Q_k(X_t,t) \mid X_s] = Q_k(X_s,s), \qquad 0 \le s \le t.$$

This furnishes the connection between Sheffer (resp., Lévy–Meixner) systems and condition (3.3). There are five Lévy–Meixner systems constructed from Hermite polynomials, Charlier polynomials $C_k(x,\mu)$, Laguerre polynomials

$L_k^{(\alpha)}(x)$, Meixner polynomials $M_k(x;\mu,q)$ and Meixner–Pollaczek polynomials $P_k(x;\mu,\zeta)$, respectively. The resulting Lévy processes X_t are standard Brownian motion B_t, the standard Poisson process N_t, the Gamma process G_t, the Pascal process P_t and the Meixner process H_t, respectively. See Schoutens and Teugels [21, 23] for details on all these processes and families of orthogonal polynomials.

Brownian motion is not of interest to us since the corresponding last line of Table 1 has been established by Glasserman and Yu [14]. As for the remaining four processes, in the light of condition (3.3), the martingale relations [21]

$$
\begin{aligned}
\mathbf{E}[C_k(N_t,t) \mid N_s] &= \left(\frac{s}{t}\right)^k C_k(N_s,s),\\
\mathbf{E}\big[L_k^{(t-1)}(G_t) \mid G_s\big] &= L_k^{(s-1)}(G_s),\\
\mathbf{E}[M_k(P_t;t,q) \mid P_s] &= \frac{(s)_k}{(t)_k} M_k(P_s;s,q),\\
\mathbf{E}[P_k(H_t;t,\zeta) \mid H_s] &= P_k(H_s;s,\zeta),
\end{aligned}
\tag{4.1}
$$

valid for $0<s<t$, prompt us to choose the basis functions in Table 2. [Note that $(t)_k = t(t+1)\cdots(t+k-1)$ is the Pochhammer symbol.] When specializing the bounds (3.4) and (3.5) to our examples, we will require the orthogonality properties

$$\mathbf{E}[C_k(N_t,t)C_l(N_t,t)] = t^{-k}k!\delta_{kl}, \tag{4.2}$$

$$\mathbf{E}\big[L_k^{(t)}(G_t)L_l^{(t)}(G_t)\big] = \frac{\Gamma(k+t+1)}{k!}\delta_{kl}, \tag{4.3}$$

$$\mathbf{E}[M_k(P_t;t,q)M_l(P_t;t,q)] = \frac{k!}{(t)_k q^k}\delta_{kl}, \tag{4.4}$$

$$\mathbf{E}[P_k(H_t;t,\zeta)P_l(H_t;t,\zeta)] = \frac{\Gamma(k+2t)}{(2\sin\zeta)^{2t}k!}\delta_{kl}, \tag{4.5}$$

TABLE 2
Lévy–Meixner systems

Process	**Notation**	**Basis polynomials $\psi_{nk}(x)$**	**Parameters**
Meixner	H_t	$\psi_{nk}^{\mathrm{M}}(x) = P_k(x;t_n,\zeta)$	$0<\zeta<\pi$
Standard Poisson	N_t	$\psi_{nk}^{\mathrm{P}}(x) = t_n^k C_k(x,t_n)$	
Gamma	G_t	$\psi_{nk}^{\mathrm{G}}(x) = L_k^{(t_n-1)}(x)$	
Pascal	P_t	$\psi_{nk}^{\mathrm{Pa}}(x) = (t_n)_k M_k(x;t_n,q)$	$0<q<1$

as well as a way to express the squares of the basis functions as series of basis functions. We will denote by $d_{ki}(t_n)$ the *linearization coefficients* in the expansion

$$\psi_{nk}(x)^2 = \sum_{i=0}^{2k} d_{ki}(t_n)\psi_{ni}(x). \tag{4.6}$$

Where distinction is necessary, the linearization coefficients corresponding to the four families in Table 2 will be written as $d^{\mathrm{P}}_{ki}(t_n)$, $d^{\mathrm{G}}_{ki}(t_n)$, $d^{\mathrm{Pa}}_{ki}(t_n)$ and $d^{\mathrm{M}}_{ki}(t_n)$, respectively. The same superscripts will adorn other quantities to distinguish the four cases, namely, the Meixner, Poisson, Gamma and Pascal process as in Table 2.

Among these processes, the Meixner process has the most significance in applications. Clearly, a financial model will impose geometric Meixner dynamics (as in Proposition 3) rather than the linear process which may become negative. But then a convergence analysis in the spirit of Glasserman and Yu [14] is impossible with polynomial basis functions as the geometric Meixner process does not have finite moments of all orders. Instead, we propose to use basis functions of logarithmic growth,

$$\psi^{\mathrm{M,log}}_{nk}(x) = P_k(\log x; t_n, \zeta). \tag{4.7}$$

Then our convergence result for the Meixner process (Theorem 4 below) can be applied. Similarly, models based on the geometric Poisson ([24], Section 112.7.1) or geometric Pascal processes can be reduced to the linear case by modifying their respective basis functions analogously.

5. Unbounded state space: The single-period problem.

5.1. *Main result and first steps of the proof.* We now state our main result about the single-period problem where our option has the exercise times $0 = t_0 < t_1 < t_2$. As noted above, the geometric Meixner model is contained in this result by modifying the basis functions according to (4.7).

THEOREM 4. *Suppose $m = 2$, that S_t is a Meixner process and that the basis functions are as in the first line of Table 2. Put $(u, v) = (8, 8)$. If the number N of paths and the number K of basis functions satisfy $N \geq K^{(u+\varepsilon)K}$ for some positive ε, then*

$$\lim_{N\to\infty} \sup_{|\beta|=1} \mathbf{E}[|\beta - \hat{\beta}|^2] = 0.$$

If $N \leq K^{(v-\varepsilon)K}$, then

$$\lim_{N\to\infty} \sup_{|\beta|=1} \mathbf{E}[|\beta - \hat{\beta}|^2] = \infty.$$

For the standard Poisson, Gamma and Pascal processes, with their respective basis functions from Table 2, the same holds if (u, v) is replaced by $(10, 4)$, $(8, 8)$ and $(11, 7)$, respectively.

The announced critical rate $\log N/\log\log N$ in Table 1 then follows from the fact that the solution of $N = K^{cK}$ satisfies $K \sim c^{-1}\log N/\log\log N$ (see, e.g., de Bruijn [6]).

Looking at (3.4) and (3.5) we begin the proof of Theorem 4 by bounding $\|\Psi_1\|$ and $\|\Psi_1^{-1}\|$, defined by (3.1) and Table 2. As in Section 2, the letter c denotes various positive constants whose value is irrelevant.

LEMMA 5. *As $K \to \infty$, the values $\|\Psi_1\|$ and $\|\Psi_1^{-1}\|$ grow at most exponentially in all four cases* (*Meixner, Poisson*, *Gamma and Pascal*), *except for $\|\Psi_1^{\mathrm{P}}\| \le c^K K^K$ and $\|\Psi_1^{\mathrm{Pa}}\| \le c^K K^{2K}$.*

PROOF. The estimates for the Meixner, Poisson and Pascal cases are easy consequences of the orthogonality relations (4.2)–(4.5) and Stirling's formula. It remains to deal with the Gamma case. The parameter $t-1$ in the martingale property (4.1) is not quite compatible with the orthogonality relation (4.3) of the Laguerre polynomials. But by the formula [26]

$$L_k^{(\alpha-1)}(x) = L_k^{(\alpha)}(x) - L_{k-1}^{(\alpha)}(x)$$

we obtain

$$\mathbf{E}[\psi_{1k}^{\mathrm{G}}(G_{t_1})\psi_{1l}^{\mathrm{G}}(G_{t_1})] = \begin{cases} -\binom{k+t_1}{k}, & k=l-1, \\ \dfrac{2k+t_1}{k+t_1}\binom{k+t_1}{k}, & k=l, \\ -\binom{k+t_1-1}{k-1}, & k=l+1, \\ 0, & |k-l|\ge 2, \end{cases} \tag{5.1}$$

hence, Ψ_1^{G} is tridiagonal. Since (5.1) grows only polynomially in k, it is clear that so does $\|\Psi_1^{\mathrm{G}}\|$. As for the inverse, note that Ψ_1^{G} is diagonally dominant so that it suffices to bound the diagonal elements of $(\Psi_1^{\mathrm{G}})^{-1}$ (see Nabben [19], Theorem 3.1); note that the τ_k from that theorem are all equal to 1 in our situation.) The diagonal elements e_k of $(\Psi_1^{\mathrm{G}})^{-1}$ can be computed recursively by [19]

$$e_{KK} = \frac{K}{K+t_1}\binom{K+t_1-1}{K-1}^{-1} \le c^K$$

and

$$e_{k-1,k-1} = \frac{k+t_1}{k}\left(\frac{2k+t_1}{k+t_1}e_{k,k} - e_{k+1,k+1}\right), \qquad 1\le k<K.$$

A straightforward backward induction shows that this implies

$$|e_{kk}| \le \bigl(4(t_1+1)\bigr)^{K-k+1} e_{KK}, \qquad 0\le k<K,$$

hence, $\|(\Psi_1^{\mathrm{G}})^{-1}\|$ grows at most exponentially too. $\square$

We proceed to bound the fourth order moments appearing in (3.4). Using (4.6) and the martingale relation (3.3), we obtain

$$\begin{aligned}
&\mathbf{E}[\psi_{2j}(S_{t_2})^2\psi_{1k}(S_{t_1})^2] \\
&\quad= \mathbf{E}\Biggl[\sum_{i=0}^{2j} d_{ji}(t_2)\psi_{2i}(S_{t_2}) \times \sum_{s=0}^{2k} d_{ks}(t_1)\psi_{1s}(S_{t_1})\Biggr] \\
&\quad= \sum_{i=0}^{2j}\sum_{s=0}^{2k} d_{ji}(t_2)d_{ks}(t_1)\mathbf{E}[\mathbf{E}[\psi_{2i}(S_{t_2}) \mid S_{t_1}]\psi_{1s}(S_{t_1})] \\
&\quad= \sum_{i=0}^{2j}\sum_{s=0}^{2k} d_{ji}(t_2)d_{ks}(t_1)\mathbf{E}[\psi_{1i}(S_{t_1})\psi_{1s}(S_{t_1})].
\end{aligned} \tag{5.2}$$

The linearization coefficients d_{ki} from the expansion (4.6) are well-studied objects for various families of orthogonal polynomials. They have combinatorial interpretations in terms of (generalized) derangements, rook polynomials and matching polynomials. See Zeng [27] for on overview of these properties, explicit formulas and many references. Paraphrasing some of these formulas ([27], Corollary 2) we have

$$d_{ki}^{\mathrm{P}}(t_n) = t_n^{2k-i} k!^2 i! \sum_{s\geq 0} \frac{t_n^s}{(s-k)!^2(s-i)!(2k+i-2s)!}, \tag{5.3}$$

$$d_{ki}^{\mathrm{G}}(t_n) = 2^{2k+i} k!^2 i! \sum_{s\geq 0} \frac{(t_n-1)_s}{4^s(s-k)!^2(s-i)!(2k+i-2s)!}, \tag{5.4}$$

$$d_{ki}^{\mathrm{Pa}}(t_n) = (1+q)^{2k+i} k!^2 i! \frac{(t_n)_k^2}{(t_n)_i} \sum_{s\geq 0} \frac{(t_n)_s(1+q)^{-2s}q^{-s}}{(s-k)!^2(s-i)!(2k+i-2s)!}, \tag{5.5}$$

$$d_{ki}^{\mathrm{M}}(t_n) = (-2\cot\zeta)^{2k+i} k!^2 i! \sum_{s\geq 0} \frac{(t_n)_s(1+(\cot\zeta)^{-2})^s}{4^s(s-k)!^2(s-i)!(2k+i-2s)!}. \tag{5.6}$$

Here it is understood that $1/n! = 0$ for n a negative integer, as is natural when extending the factorial by the Gamma function. Therefore, the sums in (5.3)–(5.6) run from $s = \max\{i,k\}$ to $s = k + \lfloor i/2 \rfloor$.

5.2. *Moment bounds in the Poisson case*. By (4.2), (5.2) and (5.3), the sum on the right-hand side of (3.4) can be estimated by

$$\begin{aligned}
&\sum_{j=0}^{K}\sum_{k=0}^{K} \mathbf{E}[\psi_{2j}^{\mathrm{P}}(S_{t_2})^2\psi_{1k}^{\mathrm{P}}(S_{t_1})^2] \\
&\qquad \leq c^K \sum_{j=0}^{K}\sum_{k=0}^{K}\sum_{i=0}^{2\min\{k,j\}} i!\Bigl(\sum_{s\geq 0} b_{jis}^{\mathrm{P}}\Bigr)\Bigl(\sum_{s\geq 0} b_{kis}^{\mathrm{P}}\Bigr),
\end{aligned} \tag{5.7}$$

where

$$b^{\mathrm{P}}_{kis} := \frac{k!^2 i!}{(s-k)!^2(s-i)!(2k+i-2s)!}.$$

It is easy to see that $b^{\mathrm{P}}_{k+1,i,k+l+1}/b^{\mathrm{P}}_{k,i,k+l} > 1$ for $i \geq 1$, $0 \leq l \leq i/2$ and $k \geq i-l$, hence, $b^{\mathrm{P}}_{k,i,k+l}$ increases in k under these conditions. From this we deduce that the s-sums in (5.7) increase in j, respectively, k:

$$\begin{aligned}\sum_{s=\max\{i,k\}}^{k+\lfloor i/2\rfloor} b^{\mathrm{P}}_{kis} &= \sum_{l=\max\{i-k,0\}}^{\lfloor i/2\rfloor} b^{\mathrm{P}}_{k,i,k+l}\\ &\leq \sum_{l=\max\{i-k,0\}}^{\lfloor i/2\rfloor} b^{\mathrm{P}}_{k+1,i,k+l+1}\\ &= \sum_{s=\max\{i,k\}+1}^{k+\lfloor i/2\rfloor+1} b^{\mathrm{P}}_{k+1,i,s}\\ &\leq \sum_{s=\max\{i,k+1\}}^{k+\lfloor i/2\rfloor+1} b^{\mathrm{P}}_{k+1,i,s}.\end{aligned}$$

Using this in (5.7) yields (recall that c may change its value in each occurrence)

$$\begin{aligned}&\sum_{j=0}^{K}\sum_{k=0}^{K} \mathbf{E}[\psi^{\mathrm{P}}_{2j}(S_{t_2})^2\psi^{\mathrm{P}}_{1k}(S_{t_1})^2]\\ &\quad\leq c^K K!^4 \sum_{i=0}^{2K}\left(\sum_{s=K}^{K+\lfloor i/2\rfloor}\frac{i!^{3/2}}{(s-K)!^2(s-i)!(2K+i-2s)!}\right)^2.\end{aligned}\tag{5.8}$$

It is plain that the summand increases in i for $K \geq 0$, $0 \leq i \leq K$ and $K \leq s \leq K + i/2$. Hence, we find that the portion $\sum_{i=0}^{K}$ of the i-sum in (5.8) can be bounded from above by

$$\begin{aligned}&(K+1)K!^3\left(\sum_{s=K}^{\lfloor 3K/2\rfloor}\frac{1}{(s-K)!^3(3K-2s)!}\right)^2\\ &\quad\leq c^K K^{5K}.\end{aligned}\tag{5.9}$$

To see the last inequality, note that the summand in (5.9) is unimodal with mode at $s = K + K^{2/3} - \frac{4}{3}K^{1/3} + O(1)$. Estimating this maximal summand, by Stirling's formula and some easy manipulations, shows that the sum in (5.9) is smaller than

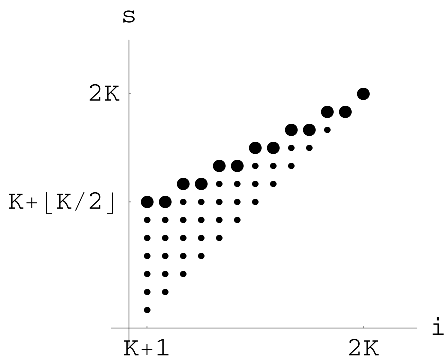

FIG. 1. *The summation range of the first sum in* (5.10).

$c^K K^K$. The remaining part $\sum_{i=K+1}^{2K}$ of the i-sum in (5.8) can be estimated by

$$\sum_{i=K+1}^{2K} i!\left(\sum_{s=i}^{K+\lfloor i/2\rfloor} \frac{i!s!}{(s-K)!^2(s-i)!(2K+i-2s)!}\right)^2$$

$$\leq c^K \sum_{i=K+1}^{2K} i!\left(\frac{i!(K+\lfloor i/2\rfloor)!}{\lfloor i/2\rfloor!^2(K+\lfloor i/2\rfloor-i)!(i-2\lfloor i/2\rfloor)!}\right)^2 \tag{5.10}$$

$$\leq c^K \sum_{i=K+1}^{2K} \frac{i!(K+\lfloor i/2\rfloor)!^2}{(K+\lfloor i/2\rfloor-i)!^2} \leq c^K K^{6K}.$$

Note that in the first line we have introduced the new factor $s!$ in the numerator. This makes the summand increasing w.r.t. the substitution $i \to i+1$, $s \to s+1$. Hence, it suffices to keep only the summands of the s-sum with $s = K + \lfloor i/2\rfloor$ (the thick dots in Figure 1) which shows the first inequality. As for the second inequality, note that the factor $i!/\lfloor i/2\rfloor!^2$ of the summand grows only exponentially and that the factor $(i-2\lfloor i/2\rfloor)!$ in the denominator is clearly negligible. Finally, the last sum in (5.10) has increasing summands which, together with Stirling's formula, implies the last inequality. By (5.8), the estimates (5.9) and (5.10) show that

$$\sum_{j=0}^{K}\sum_{k=0}^{K} \mathbf{E}[\psi_{2j}^{\mathrm{P}}(S_{t_2})^2\psi_{1k}^{\mathrm{P}}(S_{t_1})^2] \leq c^K K^{10K}.$$

In light of (3.4) and Lemma 5, the value $u = 10$ for the Poisson process in Theorem 4 is established.

As for the second assertion about the Poisson process in Theorem 4, note that, from (5.2),

$$\mathbf{E}[\psi_{2K}(S_{t_2})^2\psi_{1k}(S_{t_1})^2] = \sum_{i=0}^{2K}\sum_{s=0}^{2k} d_{Ki}(t_2)d_{ks}(t_1)\mathbf{E}[\psi_{1i}(S_{t_1})\psi_{1s}(S_{t_1})].$$

The orthogonality property (4.2) and formula (5.3) yield

$$\begin{aligned}\sum_{k=0}^{K}\mathbf{E}[\psi_{2K}^{\mathrm{P}}(S_{t_2})^2\psi_{1k}^{\mathrm{P}}(S_{t_1})^2] &\geq c^K\sum_{k=0}^{K}\sum_{i=0}^{2k} d_{Ki}^{\mathrm{P}}(t_2)d_{ki}^{\mathrm{P}}(t_1)i!\\ &\geq c^K d_{K,2K}^{\mathrm{P}}(t_2)d_{K,2K}^{\mathrm{P}}(t_1)(2K)!\\ &\geq c^K(2K)!^3\geq c^K K^{6K}.\end{aligned}$$

The second inequality follows from retaining only the summand $k=K$, $i=2K$. This makes the sum in (5.3) collapse to the summand $s=2K$, hence, the third inequality. Appealing to (3.5) and Lemma 5 completes the proof of the Poisson part of Theorem 4. Note that the preceding estimates can presumably be improved. This seems not worthwhile though; since our estimate of $\|\Psi_1^{\mathrm{P}}\|$ in Lemma 5 is sharp, we will not obtain equal values $u=v$ in Theorem 4 anyway, unless at least one of the bounds (3.4) and (3.5) was improved too.

5.3. *Moment bounds in the Meixner case.* The proofs in the remaining three cases are very similar to the Poisson case. In the Meixner case, we have

$$\begin{aligned}&\sum_{j=0}^{K}\sum_{k=0}^{K}\mathbf{E}[\psi_{2j}^{\mathrm{M}}(S_{t_2})^2\psi_{1k}^{\mathrm{M}}(S_{t_1})^2]\\ &\qquad\leq c^K\sum_{j=0}^{K}\sum_{k=0}^{K}\sum_{i=0}^{2\min\{k,j\}} i!^2\Big(\sum_{s\geq 0} b_{jis}^{\mathrm{M}}\Big)\Big(\sum_{s\geq 0} b_{kis}^{\mathrm{M}}\Big),\end{aligned}\tag{5.11}$$

where

$$b_{kis}^{\mathrm{M}} := \frac{k!^2 s!}{(s-k)!^2(s-i)!(2k+i-2s)!}.$$

Again, $b_{k,i,k+l}^{\mathrm{M}}$ increases in k and the remaining steps to show the upper bound are completely analogous to the Poisson case. This time the numerator factor $s!$ in the analogue of (5.10) appears naturally and is not introduced artificially to force some monotonicity. Moreover, the lower bound uses the same summands as in the Poisson case. Both resulting bounds are of the form $c^K K^{8K}$, hence, $u=v=8$ in Theorem 4.

5.4. *Moment bounds in the Pascal case.* We can reuse the values b_{kis}^{M} and the estimate that we just sketched:

$$\begin{aligned}&\sum_{j=0}^{K}\sum_{k=0}^{K}\mathbf{E}[\psi_{2j}^{\mathrm{Pa}}(S_{t_2})^2\psi_{1k}^{\mathrm{Pa}}(S_{t_1})^2]\\ &\qquad\leq c^K\sum_{j=0}^{K}\sum_{k=0}^{K}\sum_{i=0}^{2\min\{k,j\}} i!^3 k!\Big(\sum_{s\geq 0} b_{jis}^{\mathrm{M}}\Big)\Big(\sum_{s\geq 0} b_{kis}^{\mathrm{M}}\Big)\end{aligned}$$

$$\le c^K K!(2K)! \sum_{j=0}^{K} \sum_{k=0}^{K} \sum_{i=0}^{2\min\{k,j\}} i!^2 \Bigl(\sum_{s\ge 0} b^{\mathrm{M}}_{jis}\Bigr)\Bigl(\sum_{s\ge 0} b^{\mathrm{M}}_{kis}\Bigr)$$

$$\le c^K K!(2K)! K^{8K} \le c^K K^{11K}.$$

The lower bound poses no new difficulties either.

5.5. *Moment bounds in the Gamma case*. This part is only slightly more involved. Due to (5.1), we have three i-sums instead of one in the analogue of (5.8). The right-hand side of (5.1) can be replaced by c^K in each of these. Then one of the three i-sums equals the i-sum in (5.11) and the other two differ only in an index shift $b^{\mathrm{M}}_{k,i\pm 1,s}$ which can be easily bounded by polynomial factors. Thus the resulting growth rate is $c^K K^{8K}$, as for the Meixner case. The proof of Theorem 4 is complete.

5.6. *Side remark*: *The Bachelier model*. We finish this section with a remark about Brownian motion. If this is the underlying process S_t, then appropriate basis functions can be built from Hermite polynomials in such a way that $\|\Psi_1\|$, $\|\Psi_1^{-1}\|$ and the analogue of (4.2) grow only exponentially [14]. This is in line with the corresponding growth orders in the Gamma and Meixner cases (and in the Poisson and Pascal cases, if we renormalize our basis functions there by $1/\sqrt{k!}$ and $1/k!$, resp.). What makes the Gaussian case peculiar is that the linearization coefficients of the Hermite polynomials induce only exponential growth too when plugged into (5.2), whereas the linearization coefficients in the four cases we treat in this paper grow faster.

6. Unbounded state space: The multi-period problem. In this section we extend the main result of the preceding section (Theorem 4) to the multi-period problem, that is, to $m+1$ exercise dates $0 = t_0 < \cdots < t_m$. We know from the single-period problem that the critical rate cannot be larger than $\log N/\log\log N$, so we will be done if we can show that there is an upper bound for the mean square error of the form K^{cK} for some positive c. Fortunately, this can be deduced with little effort from a result of Glasserman and Yu [14] and the estimates from the preceding section about the single-period problem. Following [14], we assume that a representation analogous to (3.2) holds at time t_m and that the payoff functions do not grow too fast in the following sense.

THEOREM 6. *Suppose that the payoff functions satisfy the growth constraint*

$$\mathbf{E}[h_n(S_{t_n})^4] \le \max_{\nu}\left(\frac{t_{\nu+1}}{t_\nu}\right)^{2K} \max_{\nu,k} \mathbf{E}[\psi_{\nu k}(S_{t_\nu})^4], \qquad 0 \le n \le m.$$

Then the mean square error of the estimated coefficients satisfies

$$\sup_{|\beta_{m-1}|=1} \mathbf{E}[|\beta_n - \hat\beta_n|^2] \le N^{-1} c^K K^{(m-n+1)uK}, \qquad 1 \le n < m,$$

where u takes on the same values as in Theorem 4, that is, $8, 10, 8, 11$ *for* S_t, *the Meixner, standard Poisson, Gamma and Pascal process, respectively.*

PROOF. By results of Glasserman and Yu [14], Theorem 3 and the last formula before (18) on page 2096 and Jensen's inequality, we have

$$\begin{aligned}
&\sup_{|\beta_{m-1}|=1} \mathbf{E}[|\beta_n - \hat{\beta}_n|^2] \\
&\quad \le \frac{c^K}{N} \max_{1\le \nu<m} \|\Psi_\nu^{-1}\|^3 \max_{\nu,k} \mathbf{E}[\psi_{\nu k}(S_{t_\nu})^4]^{m-n} \max_{\nu,k} \mathbf{E}[\psi_{\nu k}(S_{t_\nu})^2]^2 \\
&\quad \le \frac{c^K}{N} \max_{1\le \nu<m} \|\Psi_\nu^{-1}\|^3 \max_{\nu,k} \mathbf{E}[\psi_{\nu k}(S_{t_\nu})^4]^{m-n+1}.
\end{aligned}$$

Note that Glasserman and Yu [14] assume that the moments $\mathbf{E}[\psi_{nk}(S_{t_\nu})^2]$ and $\mathbf{E}[\psi_{nk}(S_{t_\nu})^4]$ are increasing in n and k and formulate their Theorem 3 with $\mathbf{E}[\psi_{mK}^{2(4)}]$ instead of $\max_{\nu,k}\mathbf{E}[\psi_{\nu k}^{2(4)}]$. But an inspection of their proof quickly shows that taking the max in the above estimate gets rid of the monotonicity assumption. Now note that $\|\Psi_\nu^{-1}\| \le c^K$ in all our four cases by Lemma 5 and that

$$\max_{\nu,k} \mathbf{E}[\psi_{\nu k}(S_{t_\nu})^4] \le \max_\nu \sum_{j=0}^K \sum_{k=0}^K \mathbf{E}[\psi_{\nu j}(S_{t_\nu})^2 \psi_{\nu k}(S_{t_\nu})^2] \le c^K K^{uK},$$

where the double sum has been estimated in the proof of Theorem 4. □

We have thus seen that Table 1 correctly describes the general (i.e., multiperiod) situation.

7. Conclusion. Stentoft [25] and Glasserman and Yu [14] obtained apparently contradictory results about the convergence of the Longstaff–Schwartz algorithm. The main difference between their respective assumptions is the (un-)boundedness of the support of the underlying at the exercise dates. In this light the pessimistic results of Glasserman and Yu (and our Theorems 4 and 6) turn out to stem from the tails of the distribution of the underlying.

The present paper shows that Stentoft's result can be applied to Lévy models under mild assumptions and extends Glasserman and Yu's [14] results to several concrete processes. Thus we provide some evidence that Glasserman and Yu [14] were right to conjecture that their results for Brownian motion and geometric Brownian motion extend to other models.

Concerning Stentoft [25] and our Section 2: although the boundedness of the underlying induces a nice (polynomial) relation between the number of basis functions and the necessary number of Monte Carlo paths, it seems not yet completely clear that it is a harmless assumption in practice. A natural question for future research is how strongly the size of the truncation intervals influences the convergence speed of the calculated prices.

Acknowledgments. I thank Lars Stentoft, Friedrich Hubalek and an anonymous referee for helpful comments.

REFERENCES

[1] BACINELLO, A. R., BIFFIS, E. and MILLOSSOVICH, P. (2009). Pricing life insurance contracts with early exercise features. *J. Comput. Appl. Math.* **233** 27–35. MR2561492

[2] BELOMESTNY, D., BENDER, C. and SCHOENMAKERS, J. (2009). True upper bounds for Bermudan products via non-nested Monte Carlo. *Math. Finance* **19** 53–71. MR2488402

[3] BRIGO, D. and MERCURIO, F. (2006). *Interest Rate Models—Theory and Practice: With Smile, Inflation and Credit*, 2nd ed. Springer, Berlin. MR2255741

[4] CLÉMENT, E., LAMBERTON, D. and PROTTER, P. (2002). An analysis of a least squares regression method for American option pricing. *Finance Stoch.* **6** 449–471. MR1932380

[5] CONT, R., TANKOV, P. and VOLTCHKOVA, E. (2004). Option pricing models with jumps: Integro-differential equations and inverse problems. In *Proceedings of ECCOMAS* 2004 (P. Neittaanmäki et al., eds.). Jyväskylä, Finland.

[6] DE BRUIJN, N. G. (1958). *Asymptotic Methods in Analysis. Bibliotheca Mathematica* **4**. North-Holland, Amsterdam. MR0099564

[7] DE JONG, R. M. (2002). A note on "Convergence rates and asymptotic normality for series estimators": Uniform convergence rates by W. K. Newey. *J. Econometrics* **111** 1–9. MR1926535

[8] DUFRESNE, F., GERBER, H. U. and SHIU, E. S. W. (1991). Risk theory with the Gamma process. *Astin Bull.* **21** 177–192.

[9] EGLOFF, D. (2005). Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.* **15** 1396–1432. MR2134108

[10] EGLOFF, D., KOHLER, M. and TODOROVIC, N. (2007). A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *Ann. Appl. Probab.* **17** 1138–1171. MR2344302

[11] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. Wiley, New York. MR0270403

[12] FOUQUE, J.-P. and HAN, C.-H. (2009). Asymmetric variance reduction for pricing American options. In *Mathematical Modelling and Numerical Methods in Finance* (A. Bensoussan, P. Ciarlet and Q. Zhang, eds.). *Handbook of Numerical Analysis* **15** 169–187. North-Holland, Amsterdam.

[13] GERBER, H. U. and SHIU, E. S. W. (1994). Option pricing by Esscher transforms. *Transactions of Society of Actuaries* **46** 99–191.

[14] GLASSERMAN, P. and YU, B. (2004). Number of paths versus number of basis functions in American option pricing. *Ann. Appl. Probab.* **14** 2090–2119. MR2100385

[15] GODEFROY, M. (1901). *La Fonction Gamma*. Gauthier-Villars, Paris.

[16] GRIGELIONIS, B. (1999). Processes of Meixner type. *Liet. Mat. Rink.* **39** 40–51. MR1711971

[17] LONGSTAFF, F. A. and SCHWARTZ, E. S. (2001). Valuing American options by simulation: A simple least-squares approach. *Rev. Financial Stud.* **14** 113–148.

[18] MEIXNER, J. (1934). Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzeugenden Funktion. *J. London Math. Soc.* **9** 6–13.

[19] NABBEN, R. (1999). Two-sided bounds on the inverses of diagonally dominant tridiagonal matrices. *Linear Algebra Appl.* **287** 289–305. Special issue celebrating the 60th birthday of Ludwig Elsner. MR1662873

[20] NEWEY, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147–168. MR1457700

[21] SCHOUTENS, W. (2000). *Stochastic Processes and Orthogonal Polynomials. Lecture Notes in Statistics* **146**. Springer, New York. MR1761401

[22] SCHOUTENS, W. (2002). Meixner processes: Theory and applications in finance. EURANDOM Report 2002-004, EURANDOM, Eindhoven.

[23] SCHOUTENS, W. and TEUGELS, J. L. (1998). Lévy processes, polynomials and martingales. *Comm. Statist. Stochastic Models* **14** 335–349. Special issue in honor of Marcel F. Neuts. MR1617536

[24] SHREVE, S. E. (2004). *Stochastic Calculus for Finance. II. Continuous-Time Models*. Springer, New York. MR2057928

[25] STENTOFT, L. (2004). Convergence of the least squares Monte Carlo approach to American option valuation. *Manag. Sci.* **50** 1193–1203.

[26] SZEGŐ, G. (1975). *Orthogonal Polynomials*, 4th ed. *American Mathematical Society, Colloquium Publications* **XXIII**. Amer. Math. Soc., Providence, RI. MR0372517

[27] ZENG, J. (1992). Weighted derangements and the linearization coefficients of orthogonal Sheffer polynomials. *Proc. London Math. Soc.* (*3*) **65** 1–22. MR1162485

FINANCIAL AND ACTUARIAL MATHEMATICS
VIENNA UNIVERSITY OF TECHNOLOGY
WIEDNER HAUPTSTRASSE 8–10
A-1040, VIENNA
AUSTRIA
E-MAIL: sgerhold@fam.tuwien.ac.at
URL: http://www.fam.tuwien.ac.at/~sgerhold/