Finance and Stochastics
https://doi.org/10.1007/s00780-024-00550-4

Polynomial approximation of discounted moments

Chenyu Zhao[1] · **Misha van Beek**[2] · **Peter Spreij**[3] · **Makhtar Ba**[4]

Received: 12 October 2020 / Accepted: 20 June 2024

Abstract
We introduce an approximation strategy for the discounted moments of a stochastic process that can approximate the true moments for a large class of problems. These moments appear in pricing formulas of financial products such as bonds and credit derivatives. The approximation relies on a high-order power series expansion of the infinitesimal generator and draws parallels with the theory of polynomial processes. We demonstrate applications to bond pricing and credit derivatives. In the special cases that allow an analytical solution, the approximation error decreases to around 10 to 100 times machine precision for higher orders. When no analytical solution exists, we numerically compare the approximation with existing numerical techniques.

Keywords Markov processes · Pricing · Hedging · Short-rate models · Credit models · Generator · Resolvent

Mathematics Subject Classification 60J70 · 60J28 · 60J35 · 91G20 · 91G30 · 91G40

JEL Classification C32 · G12 · C58

✉ P. Spreij
spreij@uva.nl

C. Zhao
phillip.zhao@blackrock.com

M. van Beek
misha.vanbeek@bayesline.com

M. Ba
makhtar.ba@blackrock.com

1 25 Christopher Columbus Dr, Apt 1002 Jersey City, NJ 07302, USA

2 1702 Newkirk Ave, Apt 6B, Brooklyn, NY 11226, USA

3 Korteweg-de Vries Institute for Mathematics, University of Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands

4 3312 Hudson Avenue, Apt 2B, Union City, NJ 07087, USA

Published online: 04 November 2024

1 Introduction

For pricing and hedging applications, the interest is often in calculating the expected value of a discounted function of a stochastic process,

$$\mathbb{E}\left[e^{-\int_0^t r(X_s)\mathrm{d}s} f(X_t)\middle| X_0 = x\right],$$

where f describes the contingent claim and r is the risk-free rate. Sometimes another rate may be used for discounting, such as a hazard rate.

In several cases, this expectation has enough structure to allow analytical or semi-analytical solutions. For example, if the process (X_t) is an affine process and r is an affine function, then the Fourier transform of f can be used to compute the expectation up to an integral and the solution to a system of Riccati equations (see Duffie et al. [14]). Also, if (X_t) is a polynomial process as defined by Cuchiero et al. [13], f is a polynomial function and there is either no or constant discounting, then a simple analytical expression exists.

This paper introduces an approximation formula that may work in situations where no analytical expression can be found. The functional form of the approximation of order k is

$$\mathbb{E}\left[e^{-\int_0^t r(X_s)\mathrm{d}s} \langle \bar{f}^k, b^k(X_t)\rangle\middle| X_0 = x\right] \approx \langle e^{tA_k} \bar{f}^k, b^k(x)\rangle,$$

where $b^k(x)$ is a vector of certain basis functions evaluated at x (for now we take it as $(1, x, x^2, \ldots, x^{k-1})$ for a univariate process on $\mathbb{R}$, but multivariate cases will be considered) and $\bar{f}^k = (f_0, \ldots, f_{k-1})$ is a vector of length k so that the (standard) inner product of vectors $\langle \bar{f}^k, b^k(X_t)\rangle = \sum_{i=0}^{k-1} f_i b_i(x)$ represents a (polynomial) expression of the value of the contingent claim. The matrix A_k can be derived from the infinitesimal generator of the process and the function r. Naturally, when we are interested in the ith discounted moment, we can choose as basis vector $\bar{f}^k = e_i$ for $i = 0, \ldots, k-1$. Here e_i is the vector of length k that has 1 as the entry at the ith position, all other entries being zero. Note that the numbering starts with $i = 0$, which corresponds to the monomials x^i, also starting with $i = 0$. Other choices for the basis functions are equally well conceivable, and we in fact rely mostly on Hermite polynomials in the remainder of this paper.

We investigate two primary applications of this approximation. The first is in the calculation of bond prices in short-rate models. As the order increases, the approximation approaches machine precision, or falls within the margins of other numerical techniques when the true bond price has no closed-form expression. This is shown for Cox–Ingersoll–Ross (CIR) [12] and Black–Karasinski [7] bond prices. Figure 1 previews several orders of magnitude in performance gain over existing numerical techniques. This comparison was made on a simple CIR bond price to illustrate the convergence to the known analytical solution.

The second application is the calculation of spreads in a generalised Markovian model of credit rating migrations that we develop in an accompanying paper [6]. This model follows the setup of Lando [19] and assumes that companies migrate within a set of m ratings, e.g. {AAA, AA, ..., CCC, D}, according to a Markov chain (R_t). The Markov chain has an $m \times m$ generator matrix $Q(Y_t)$ that depends

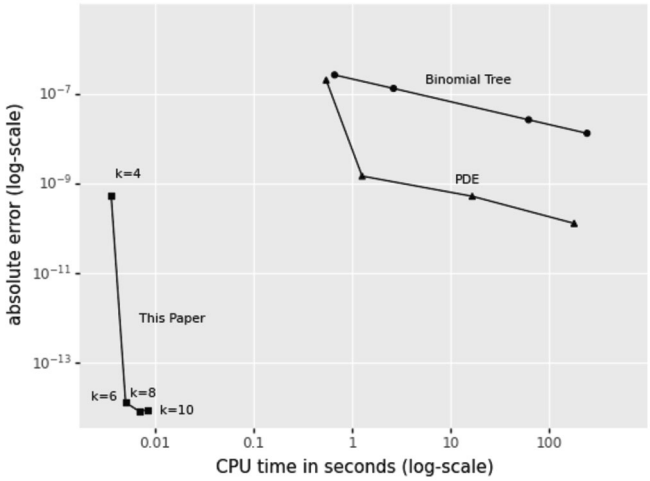

Fig. 1 Comparison of polynomial approximation against standard numerical methods for a CIR bond price. All implementations are in Python, hence not optimised for computational efficiency or memory usage

on a state Y_t. Jarrow et al. [18] derive an analytical solution to spread curves when Q is constant and does not depend on the state Y_t. Lando [19] and Arvanitis et al. [5] provide pricing equations when the generators $Q(Y_t)$ commute (i.e., $Q(y_1)Q(y_2) = Q(y_2)Q(y_1)$ for any values y_1 and y_2 that the process might take) and (Y_t) follows an affine process. Hurd and Kuznetsov [16] apply these equations to the case that $Q(y) = y_1 Q_1 + y_2 Q_2$ with (Y_t) a bivariate CIR process and Q_1 and Q_2 are two commuting generator matrices.

The commutativity requirement is highly restrictive, as pointed out by Martin [22]. There is strong empirical evidence that upgrades tend to slow down when downgrades speed up, suggesting that the upper and lower triangles of $Q(y)$ are driven separately by two negatively correlated processes. Upper and lower triangular matrices do not commute. We can use our approximation to relax the commutativity requirement as well as the CIR requirement. This relaxation allows us to cover several more stylised facts about credit migrations and spreads, as argued in our accompanying paper [6]. In the exceptional special cases where an analytical solution exists, our approximation method approaches machine precision as the order increases.

While we provide theoretical conditions for the approximation to converge, we were not able to prove that these conditions hold for any of the more interesting applications. For the CIR bond price, we come close and prove convergence up to one technical condition (see Sect. 4.1). Numerically, this condition appears to hold, but the proof remains an open problem.

In general, the approximation we propose is easy to compute. It requires application of the infinitesimal generator $\mathcal{A}$ to the terms in a polynomial basis. In the canonical univariate case, this means computing $\mathcal{A}b_0, \mathcal{A}b_1, \mathcal{A}b_2, \ldots$ for certain basis functions b_i and subsequently projecting the results on the same basis. Whereas this computation can be done by hand in many cases, it is straightforward in general for a symbolic software package.

Once the correct form of the matrix A_k is found in this way, the computation of moments is very fast. Al-Mohy and Higham [2] offer a very efficient and numerically accurate algorithm for calculating the action of the matrix exponential $e^{tA_k}\bar{f}^k$ for a series of times t. Subsequent computing of the approximation for a given state x can also be very efficient. In the univariate case, per Horner's method, this takes k additions and k multiplications, where we found that usually an order of $k = 20$ is sufficient. This is especially convenient for empirical methods such as maximum likelihood estimation, Kalman filtering or MCMC methods, where we typically distinguish between construction of $e^{tA_k}\bar{f}^k$ that happens once per likelihood evaluation, and subsequent computation of the inner product $\langle e^{tA_k}\bar{f}^k, b^k(x)\rangle$ which is required as often as once per time-step within a single likelihood evaluation.

Apart from the applications that this paper explicitly investigates, we list several other applications. First, consider the generic problem of estimating the parameters of a discretely sampled continuous-time stochastic process. Naturally, the availability of moments lends itself to generalised-method-of-moments-based estimation, such as described by Zhou [30]. But maximum likelihood estimations can also benefit from moment approximations. There is a one-to-one mapping between moments and cumulants, and given some technical conditions, probability density functions can be accurately approximated by cumulants using Gram–Charlier type A series. Such approximations can be much more efficient than PDE, tree or simulation-based approaches (see Aït-Sahalia [1]). Second, Gram–Charlier-based approximations of density functions can also be useful for option pricing; see Popovic and Goldsman [24], Tanaka et al. [28]. Specifically, Chateau and Dufresne [11] show that a European option price has a Gram–Charlier-based approximation that is linear in the moments of the process. This approximation does not, however, improve with the number of moments included. Finally, Cuchiero et al. [13] outline how variance reduction techniques can benefit from knowledge about the moments.

To the best of our knowledge, this paper is the first to use an approximation of the infinitesimal generator by applying it to a polynomial basis. There are, however, myriad other ways to compute approximations of moments of stochastic processes such as bond prices. Apart from standard PDE, tree or simulation-based approaches, we mention a few. First, Tourrucôo et al. [29] and Antonov and Spector [3] approximate solutions of bond-pricing-related PDEs using perturbation techniques. Perturbation approaches may quickly accumulate errors over time; hence this technique is not suitable for longer-maturity bond prices. Second, Stehlíková and Capriotti [27] use an exponent expansion technique (see Capriotti [10]) that yields quickly converging results at shorter maturities for Black–Karasinski bonds, but errors accumulate quickly at medium to longer maturities. However, the authors present a convolution approach to get higher accuracy at longer maturities (up to 4 significant digits in the 20-year discount factor). Finally, Li et al. [20] rely on chaos expansion techniques to obtain the moment-generating function of the integral of a mean-reverting process. This approach yields simple explicit formulas, as demonstrated by pricing a quanto CDS.

This paper is organised as follows. Section 2 sets up the general notation of the paper. Section 3 derives the theoretical results behind our approximation. Finally, Sects. 4 and 5 show applications to the aforementioned short-rate models and credit derivatives, respectively. Appendix A contains some supporting technical results.

2 General notation

This section sets up general notation for the remainder of this paper. We borrow parts of our setup from Cuchiero et al. [13]. Central in our analysis is the Feller process X, i.e., a time-homogeneous continuous-time Markov process, with state space denoted by $D \subseteq \mathbb{R}^d$. If the process X is not conservative, we augment the state space with a point $\Delta \notin D$ to get the augmented state space D_Δ. This point is usually referred to as the *cemetery state* for killed processes and is used to incorporate discounting. Any function f on D is extended to D_Δ by the convention $f(\Delta) = 0$. We further consider the Feller semigroup $(S_t)_{t\geq 0}$ (often simply denoted (S_t) or even S) given by

$$S_t f(x) := \mathbb{E}_x[f(X_t)] \tag{2.1}$$

and acting on all Borel-measurable functions $f : D_\Delta \to \mathbb{R}$ for which the expectation is well defined. Here we used $\mathbb{E}_x$ to denote expectation under the law $\mathbb{P}_x$ that is such that the process starts in $x \in D$, i.e., $\mathbb{P}_x[X_0 = x] = 1$. When we need (in applications to follow) that certain multi-powers of the X_t have a finite expectation, it will be implicitly assumed that such moments exist and are finite. We also specify later the space $\mathcal{F}$ of functions f to which the operators S_t are applied, together with the norm on it. It will be such that the S_t have finite operator norm. We denote by $\mathcal{A}$ the associated linear operator that describes the process, i.e.,

$$\mathcal{A}f(x) := \lim_{t\downarrow 0} \frac{S_t f(x) - f(x)}{t} \tag{2.2}$$

for all functions $f : D_\Delta \to \mathbb{R}$, $f \in \mathcal{F}$, for which this limit is well defined. This set is the domain of $\mathcal{A}$, denoted by $\mathcal{D}(\mathcal{A})$.

This paper relies on series representations, mostly with respect to some orthonormal basis. We start with a sequence b of linearly independent functions $b_i : D \to \mathbb{R}$; so $b = (b_i)_{i=0}^\infty$. We then further have sequences of real numbers $b(x) = (b_i(x))_{i=0}^\infty$ for $x \in D$. We denote by $\mathcal{P}$ the space of functions that can be written as a *formal (power) series* with respect to b, i.e., for all $f \in \mathcal{P}$, there exists a sequence $\bar{f} = (\bar{f}_i)_{i=0}^\infty$ in $\mathbb{R}$ such that

$$f(x) = \langle \bar{f}, b(x) \rangle := \sum_{i=0}^{\infty} \bar{f}_i b_i(x). \tag{2.3}$$

We need the sum in (2.3) to be convergent in a suitable norm. The sequence $\bar{f}$ then denotes an infinite-dimensional vector representation of the function, and $\langle \cdot, \cdot \rangle$ is the inner product notation for the infinite sum of vectors or vector-valued functions evaluated at x. We formalise this now.

Consider a separable Hilbert space of functions on D, with a certain inner product. A typical example is the L^2-space with respect to an underlying measure. As a specific example, we mention the L^2-space of (Borel-measurable) functions f on $\mathbb{R}$ satisfying $\int_{\mathbb{R}} f(x)^2 \phi(x) \mathrm{d}x < \infty$, where ϕ is the standard normal density. Clearly, this space contains all polynomials. Moreover, the Hermite polynomials form an orthonormal basis for this space, and if we choose the b_i in (2.3) as these polynomials,

the squared L^2-norm $\|f\|^2$ coincides with $\sum_{i=0}^{\infty} \bar{f}_i^2$. If we take the b_i as the monomic polynomials in (2.3), then in the same L^2-space, $\|f\|^2$ can be written as $\bar{f}^\top P \bar{f}$ for a certain strictly positive definite infinite-dimensional matrix P. In what follows, **we always assume**, unless stated otherwise, that $\mathcal{P}$ is a Hilbert space of functions with respect to an appropriate inner product $\langle \cdot, \cdot \rangle$ and that it admits an orthonormal basis b such that any $f \in \mathcal{P}$ can be represented as in (2.3) with a sum that is convergent in $\mathcal{P}$. We denote by $\mathcal{H}$ the Hilbert space (actually an ℓ^2-space) of vectors $\bar{f}$ associated with $f \in \mathcal{P}$, for which we then automatically have $\langle \bar{f}, \bar{f} \rangle < \infty$. Here, with a little ambiguity of notation, $\langle \cdot, \cdot \rangle$ denotes the inner product of ℓ^2. It follows that the ℓ^2-norm of $\bar{f}$ coincides with the Hilbert space norm of f. With a little but innocuous abuse of notation, we invariably use the same symbol to denote inner products, sums, norms and thus have $\langle \bar{f}, \bar{f} \rangle = \|\bar{f}\|^2 = \|f\|^2 = \langle f, f \rangle$ and $f = \langle \bar{f}, b \rangle$.

Of course, other L^2-spaces with respect to a different weight function than ϕ and other orthogonal polynomials are also conceivable. For instance, if the relevant domain is $[0, \infty)$, there are clearly relations to finance as prices are nonnegative, and Laguerre polynomials come into the picture. If the domain of the functions is a compact interval like $[-1, 1]$, then Jacobi, Chebychev or Legendre polynomials are relevant to consider. Moreover, in the case of such a compact domain, also Fourier expansions are a possibility, whereas for functions defined on all of $\mathbb{R}$, one may also think of expansions on a basis of Haar functions. Of course, one has to ensure that the S_t form a semigroup of bounded operators on the underlying function space.

By $\mathcal{H}_k$, we denote the subspace of $\mathcal{H}$ of sequences $\bar{f}$ with $\bar{f}_i = 0$ for all $i \geq k$. We further let $\mathcal{P}_k$ be the space of 'polynomials' with k terms (the terminology is suggestive) associated with $\mathcal{H}_k$, meaning that $f \in \mathcal{P}_k$ if and only if $f = \sum_{i=0}^{k-1} \bar{f}_i b_i$. An element $\bar{f}$ of $\mathcal{H}_k$ is also in a natural way identified with a vector $(f_0, \ldots, f_{k-1}) \in \mathbb{R}^k$. Note that $(a, b, c, \ldots)$ for vectors or scalars a, b and c denotes vertical stacking into one (potentially infinitely) long vector. This convention is used **throughout the paper**. In the case of Hermite polynomials of a single variable, the space $\mathcal{P}_k$ consists of all polynomials of degree $k-1$ or less. We denote the Hermite polynomial of order k by $\mathrm{He}_k(x)$. For higher-dimensional state spaces $D \subseteq \mathbb{R}^n$, we use multi-index notation k to denote $\mathrm{He}_k(x) = \mathrm{He}_{k_1}(x_1)\mathrm{He}_{k_2}(x_2) \cdots \mathrm{He}_{k_n}(x_n)$ for $k_i \in \mathbb{N}_0$, $\sum_{i=1}^n k_i = k$. For example, for $n = 2$, $\mathrm{He}_0(x) = 1$ is scalar, $\mathrm{He}_1(x) = (\mathrm{He}_1(x_1), \mathrm{He}_1(x_2))$ is a 2-vector, $\mathrm{He}_2(x) = (\mathrm{He}_2(x_1), \mathrm{He}_1(x_1)\mathrm{He}_1(x_2), \mathrm{He}_2(x_2))$ is a 3-vector, etc. Then under this notation, for $D = \mathbb{R}^n$ and $D = \mathbb{R}^n_+$, $b(x) = (\mathrm{He}_0(x), \mathrm{He}_1(x), \mathrm{He}_2(x), \ldots)$ has the same symbolic representation as the one-dimensional case. Another useful state space to which we return later is the set of basis vectors $D = \{e_1, \ldots, e_d\}$ of $\mathbb{R}^d$. In this case, $b(x) = \mathrm{He}_1(x)$ is an adequate basis as other powers (i.e., $\mathrm{He}_k(x)$ in the vector sense and with $k \neq 1$) of unit vectors are linearly dependent. A general notational convention follows. We write $b^k(x) = (b_0(x), \ldots, b_{k-1}(x))$, which is also identified with $b^k(x) = (b_0(x), \ldots, b_{k-1}(x), 0, \ldots)$, the vector where the first k entries of $b(x)$ are followed by zeros.

The crux of our approximation theory relies on finite-dimensional modifications of mappings on $\mathcal{H}$. To this end, we introduce some notation that involve projections and subspaces. Let $P_k : \mathcal{H} \to \mathcal{H}_k$, $k \geq 1$, be a sequence of projection operators, i.e., idempotent operators with co-domain $\mathcal{H}_k$. Above we have made the special choice where $\mathcal{H}$ is the ℓ^2-space of sequences $\bar{f}$ (satisfying $\langle \bar{f}, \bar{f} \rangle < \infty$), and $\mathcal{H}_k$ is the space of finite vectors $f^k = (f_0, \ldots, f_{k-1})$ also identified with $(f_0, \ldots, f_{k-1}, 0, \ldots)$.

While multiple projections can be considered, this paper assumes that P_k is the *finite-section* projection, i.e., we take the Hilbert space of sequences in ℓ^2 with $P_k \bar{f} = (\bar{f}_0, \ldots, \bar{f}_{k-1}, 0, \ldots)$. Correspondingly, if we fix a sequence of basis functions $b(x)$ in $\mathcal{P}$ that forms an orthonormal basis and let P_k be the orthogonal projection on $\mathcal{P}_k$ which is the linear span of $b_0, \ldots, b_{k-1}$, then $P_k f \in \mathcal{P}_k$ has the representation $P_k \bar{f} \in \mathcal{H}_k$. Here we deliberately use the same notation P_k for the projections onto $\mathcal{P}_k$ and $\mathcal{H}_k$.

In general, we consider a fixed sequence of basis functions $b(x)$ that forms an orthonormal basis, and projections $P_k : \mathcal{H} \to \mathcal{H}_k$ onto finite-dimensional subspaces $\mathcal{H}_k$. Note that the orthogonal projections $P_k : \mathcal{H} \to \mathcal{H}_k$ given by $\bar{f}^k := P_k \bar{f} \in \mathcal{H}_k$ have the nice property that the operator norm $\|P_k\| = 1$ and $\|\bar{f}_k\| \leq \|\bar{f}\|$. For any linear operator $B : \mathcal{H} \to \mathcal{H}$, define the *finite-section approximation*

$$B_k := P_k B|_{\mathcal{H}_k} : \mathcal{H}_k \to \mathcal{H}_k. \tag{2.4}$$

The restriction to $\mathcal{H}_k$ lets us interpret B_k as a $k \times k$ matrix when $\mathcal{H}_k$ has dimension k, as in the case that we just considered. Note that the B_k are automatically bounded operators, whereas B is typically only closed in the cases that are of interest for us. Where necessary, one can also consider the B_k as operators $B_k := P_k B : \mathcal{H} \to \mathcal{H}_k$ simply by embedding $\mathcal{H}_k$ into $\mathcal{H}$. This abuse of interpretation should not cause any confusion, and we freely switch between B_k defined on $\mathcal{H}_k$ and on $\mathcal{H}$.

If B is invertible, we can find a unique solution $\bar{g} \in \mathcal{H}$ to the system $B\bar{g} = \bar{f}$, namely $\bar{g} = B^{-1}\bar{f}$. The approximation by the finite-dimensional system $B_k \bar{g}^k = \bar{f}^k$ is called the *finite-section method* (FSM). The main theorem in Sect. 3 applies the FSM to the resolvent of a map $A : \mathcal{H} \to \mathcal{H}$, i.e., if we let $B = \lambda I - A$, then $B^{-1} = R(\lambda, A) := (\lambda I - A)^{-1}$. We use the same notation for the resolvent of a linear operator $\mathcal{A}$, namely $R(\lambda, \mathcal{A}) := (\lambda I - \mathcal{A})^{-1}$. The resolvent is said to be *defined* if λ is in the resolvent set, i.e., if the inverse exists.

Finally, some more notational conventions follow. On finite-dimensional spaces, we use the notation 0, I and e_i to represent the zero matrix, the identity matrix or operator and the ith standard basis vector (the ith column of I), respectively. For these objects, we assume that the size is clear from the context and may be infinite, unless explicitly provided. The operator $\otimes$ stands for the Kronecker or matrix direct product, and $\mathrm{diag}(a)$ represents the diagonal matrix with the elements of a vector a on the diagonal. We also write $\mathrm{diag}_i(a_i)$, where i identifies all elements of the vector a.

3 Polynomial moment approximation theory

This section contains the heart of this paper, i.e., the theoretical results behind the polynomial approximation that we propose. Consider a Feller process X on a state space D with Feller semigroup of operators (S_t) and (infinitesimal) generator $\mathcal{A}$.

Definition 3.1 A Feller semigroup $S = (S_t)$ and the associated Feller process X is called *f-sequential* if for $f \in \mathcal{P}$ and $t \geq 0$, (i) $S_t f$ is well defined and (ii) $S_t f \in \mathcal{P}$. Furthermore, (S_t) is called *sequential* if it is *f-sequential* for all $f \in \mathcal{P}$.

Remark 3.2 Cuchiero et al. [13] call a time-homogeneous Markov process X polynomial with semigroup (S_t) if for all $k \geq 0$, we have $S_t f \in \mathcal{P}_k$ for all $f \in \mathcal{P}_k$ and $t \geq 0$. The key difference with sequential processes is the additional requirement that a polynomial of order k cannot return a higher order polynomial when (S_t) is applied. Therefore all polynomial processes are sequential.

Remark 3.3 The quantity $S_t f$ appears to represent a standard moment in (2.1), but can also represent a discounted moment when (S_t) is not a conservative semigroup, namely through appropriate specification of the killing rate at which the process jumps to the cemetery state Δ. For more details, see Sect. 4 or Duffie et al. [14, Sect. 11]. This feature allows us to show that the Cox–Ingersoll–Ross (CIR) bond price is an expectation of an f-sequential process with killing rate (see Sect. 4.1). Due to the stochastic killing rate, this process is not polynomial. We should note that we were not able to verify all technical conditions necessary to apply our polynomial moment approximation theory to the CIR bond price. One technical condition remains an open problem (see Sect. 4.1).

Remark 3.4 The time-homogeneity of the Feller assumption can potentially be relaxed to piecewise time-homogeneity. A typical example happens in the context of local models where up to a time τ_1, the process X evolves according to a certain semigroup and starting from τ_1 according to another semigroup, and then repeatedly changes at times τ_k. These times are usually chosen to correspond to tenors of derivatives. For some practical applications, this is useful, but it complicates notation and analysis considerably, and is not pursued further in the present paper.

Any sequential semigroup is a family (S_t) of linear maps from $\mathcal{P}$ to $\mathcal{P}$, and hence with a fixed basis b, these induce linear maps $\bar{S}_t$ from the sequence space $\mathcal{H}$ to $\mathcal{H}$ which have an infinite-dimensional matrix representation. Let $g(t) := S_t f$. Then, using the vector representations $\bar{g}(t)$ of $g(t)$ and $\bar{f}$ of f, we may write

$$\bar{g}(t) = \bar{S}_t \bar{f},$$

where the ji-element $\bar{S}_{t,ji}$ is defined as $\bar{S}_{t,ji} = \bar{c}_j^{(i)}(t)$ resulting from the representation of $c^{(i)}(t) := S_t b_i$.

In an analogous way, we consider the derivative in (2.2). Assuming that each b_i belongs to $\mathcal{D}(\mathcal{A})$, we put $c^{(i)} := \mathcal{A} b_i$ and then the $c^{(i)}$ belong to $\mathcal{P}$ as well, i.e., $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{P}$ (which notably does *not* imply that $\mathcal{D}(\mathcal{A}) = \mathcal{P}$). In all examples that follow, this assumption is satisfied. As all the $c^{(i)}$ belong to $\mathcal{P}$, we can represent them by their coordinate vectors $\bar{c}^{(i)} \in \mathcal{H}$ with elements denoted by $\bar{c}_j^{(i)}$. We then define the infinite-dimensional matrix A representing a map from $\mathcal{H}$ into $\mathcal{H}$ having ji-entry $A_{ji} = \bar{c}_j^{(i)}$. We call A the *matrix generator* of the process X. In fact, any linear map, call it $\mathcal{A}$ again, from $\mathcal{P}$ into itself naturally induces a map $A : \mathcal{H} \to \mathcal{H}$ in a similar way. As any $f \in \mathcal{P}$ can be identified with a sequence $\bar{f} \in \mathcal{H}$, and similarly a function $g \in \mathcal{P}$ can be identified with a sequence $\bar{g} \in \mathcal{H}$, one can define $\bar{g} = A\bar{f}$ if $g = \mathcal{A} f$ for $f \in \mathcal{D}(\mathcal{A})$.

Since a generator $\mathcal{A}$ of a semigroup is a closed operator, so is A. To see this, we use the duality between elements of $\mathcal{P}$ and those of $\mathcal{H}$. We use that $\mathcal{A}$ is closed if and only

if $\mathcal{D}(\mathcal{A})$ is complete with respect to the graph norm given by $\|f\|_{\mathcal{A}}^2 = \|f\|^2 + \|\mathcal{A}f\|^2$ (see Bobrowski [8, Exercise 7.3.3]) and similarly that A is closed if and only if $\mathcal{D}(A)$ is complete with respect to the graph norm given by $\|\bar{f}\|_A^2 = \|\bar{f}\|^2 + \|A\bar{f}\|^2$. But by construction, $\|f\|_{\mathcal{A}} = \|\bar{f}\|_A$. In the sequel, we freely switch between $f \in \mathcal{P}$ having an orthogonal expansion in terms of a sequence $\bar{f}$, and between $\mathcal{A}f$ and $A\bar{f}$.

For Feller semigroups, we have $\partial_t S_t f = \mathcal{A}S_t f$ for $f \in \mathcal{D}(\mathcal{A})$. Hence for $g(t) = S_t f$, we have $\partial_t g(t) = \mathcal{A}g(t)$, and in the corresponding sequence space $\mathcal{H}$, one has $\partial_t \bar{g}(t) = A\bar{g}(t)$. Paralleling finite-dimensional notation, we write $\bar{g}(t) = e^{tA}\bar{f}$ as is done for polynomial processes in Cuchiero et al. [13, Theorem 2.7], although in general the matrix A is genuinely infinite-dimensional (and has infinite norm, so that we should only interpret this exponential action as a definition, not as a method of computing $\bar{g}(t)$).

We use the finite-dimensional matrix A_k to approximate the semigroup. That is, in ordinary finite-dimensional notation, we use

$$\bar{g}^k(t) := \bar{S}_t^k \bar{f}^k, \qquad \bar{S}_t^k := e^{tA_k}, \tag{3.1}$$

to approximate $\bar{g}(t)$. We use the name *polynomial approximation* as a consequence of the polynomial structure of the approximating $g^k(t)$ as functions in x, when the A_k are taken as in (2.4). Our main theoretical result is on the convergence of the approximation in (3.1). Before we state it, we need one more definition, for which we use the convention that elements of $\mathcal{P}_k$ are considered as elements of $\mathcal{P}$ by the natural embedding.

Definition 3.5 Consider a closed and invertible linear operator $B : \mathcal{H} \to \mathcal{H}$ and its finite-section approximation B_k as in (2.4). The FSM is said to be

1) f*-applicable* to B if $\lim_{k\to\infty} B_k^{-1}\bar{f}^k = B^{-1}\bar{f}$, with $f \in \mathcal{P}$;
2) *applicable* to B if the FSM is f-applicable for all $f \in \mathcal{P}$.

We often need that the FSM is applicable (or f-applicable) to $B = \lambda I - A$, which is the inverse of the resolvent. In these cases, we implicitly assume that λ is chosen in the resolvent set, as this is always possible for sufficiently large λ by the properties of the resolvent. Note also that this definition hinges on the fact that finite sections A_k of A trivially extend to finite sections $(\lambda I - A)_k = \lambda I - A_k$ of $\lambda I - A$, and that $\lambda I - A$ is closed since A is closed.

Remark 3.6 The applicability of the FSM exists with different notions of convergence. Hagen et al. [15, Sect. 1.1.1.] consider *strong* convergence of operators, whereas Lindner [21, Sect. 2.6.3] uses the weaker concept of *strict* convergence. We opt for strong convergence. Recall that this means the following, in a more general situation than ours: If $T_n, T : X \to X$ are operators on a Banach space X with a common domain $\mathcal{D} \subseteq X$, then the T_n strongly converge to T if $\|T_n x - Tx\| \to 0$ for all $x \in X$. For completeness, we also mention a characterisation of strict convergence of u_n to u in ℓ^2: This takes place if the norms $\|u_n\|_2$ are bounded and if the u_n converge to u element-wise. It is evident that this is a weaker concept.

We recall the following result. For a strongly continuous semigroup (S_t) acting on a Banach space $\mathcal{P}$ with generator $\mathcal{A}$ (having domain $\mathcal{D}(\mathcal{A})$), there exist $C \geq 1$ and $w \geq 0$ such that the operator norm satisfies

$$\|S_t\| \leq Ce^{wt}; \tag{3.2}$$

see Bobrowski [8, Eq. (7.14)].

In the main result in Theorem 3.7 below, we assume that the A_k are derived from the given generator A by projections. Let P_k be projections of $\mathcal{P}$ onto subspaces $\mathcal{P}_k$ with norm $\|P_k\| \leq 1$. Let A_k be the finite-section approximation of A as in (2.4). The A_k can also be seen as operators (with domains $\mathcal{D}(A_k)$) on $\mathcal{P}_k$ by identifying A_k with $P_k A P_k$. Similarly, for $f^k \in \mathcal{D}(A_k)$, we consider $R(\lambda, A_k)\bar{f}^k \in \mathcal{P}_k$ as an element of $\mathcal{P}$. For the proof of the theorem, it would be convenient that the same bound (3.2) applies to the semigroups S^k uniformly in k. The condition that the (S_t^k) all satisfy (3.2) is at first glance reasonable. Indeed, for the easy case that A is a bounded operator, it is almost trivially satisfied as then $\|A_k\| \leq \|A\|$ and $\|S_t^k\| \leq e^{\|A_k\|t} \leq e^{\|A\|t}$. However, in many relevant situations, boundedness of A appears to be too restrictive. Therefore we make the weaker assumption that for a semigroup (S_t) and all bounded and continuous f, there exist nonnegative constants C_f and w_f such that

$$\|S_t f\| \leq C_f e^{w_f t} \qquad \text{for all } k \geq 1, t \geq 0. \tag{3.3}$$

Note that here we do not require the semigroup to be strongly continuous, as this would lead to the stronger property (3.2). We impose this weaker assumption on the semigroups (S_t^k), but uniformly in k, in Theorem 3.7. In that result, we also need one extra assumption on the projection f^k of f, namely that for all *relevant* (depending on the application at hand) functions f and all $f^k = \mathcal{P}_k f$, it holds that

$$C_f^u := \sup_k C_{f^k} \quad \text{and} \quad w_f^u := \sup_k w_{f^k} \qquad \text{are finite.} \tag{3.4}$$

Note that both finite suprema in principle depend on the particular function f at hand. A similar condition applies to $\bar{f}^k$. Namely, if (3.3) and (3.4) hold, then we have

$$\|R(\lambda, A_k, f)\| = \left\| \int_0^\infty e^{-\lambda t} \bar{S}_t^k f \, \mathrm{d}t \right\| \leq \frac{C_f^u}{\lambda - w_f^u}$$

for $\lambda \geq w_f^u$ and

$$\sup_k \|R(\lambda, A_k, f^k)\| \leq \frac{C_f^u}{\lambda - w_f^u} < \infty$$

for $\lambda \geq w_f^u$. To avoid complications in its proof, we impose in Theorem 3.7 some restricting assumptions like (3.6) below. These restricting assumptions are immediately satisfied if

$$\bar{f}^k = \bar{f} \qquad \text{for all sufficiently large } k. \tag{3.5}$$

Clearly, this last strong assumption does not hold in general, but is satisfied in important applications like that with the CIR model in Sect. 4.1. But (3.6) below of course holds if the sequence $(A\bar{f}^k)$ converges. Moreover, in that case, because A is a closed operator, it holds that necessarily $A\bar{f}^k \to A\bar{f}$.

Theorem 3.7 *Consider an f-sequential process for some $f \in \mathcal{P}$. Assume (S_t) satisfies (3.2) and in addition that the generators A_k are such that the generated semigroups (S_t^k) all satisfy the same norm bound (3.3), i.e., $\|S_t^k f\| \le C_f e^{w_f t}$ for all $k \ge 1, t \ge 0$. Assume also (3.4). Let f be such that also*

$$\alpha := \sup_k \|A\bar{f}^k\| < \infty. \tag{3.6}$$

If the FSM is f-applicable to $R(\lambda, A)^{-1} = \lambda I - A$ for all big enough λ, then $\bar{g}^k(t) \to \bar{g}(t)$ (i.e., $e^{tA_k}\bar{f}^k \to e^{tA}\bar{f}$) as $k \to \infty$, with convergence in the ℓ^2-norm.

Proof Let $t > 0$ as the case $t = 0$ is trivial. Consider the Phragmén representation of the semigroup (S_t); see Neubrander [23] and Lemma A.1. For $f \in \mathcal{D}(\mathcal{A})$, it holds that $S_t f = \lim_{\lambda\to\infty} S_t(\lambda, \mathcal{A}, f)$ with, see (A.2),

$$S_t(\lambda, \mathcal{A}, f) = \lambda \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{(n-1)!} e^{n\lambda t} R(n\lambda, \mathcal{A}) f,$$

where $R(\lambda, \mathcal{A}) := (\lambda I - \mathcal{A})^{-1}$ denotes the resolvent. Naturally, when switching from $\mathcal{P}$ to $\mathcal{H}$, we can write this in matrix form with $g(t) = S_t f$ and $\bar{g}(t) = \bar{S}_t \bar{f} = e^{At}\bar{f}$. One has, see also Corollary A.2,

$$\bar{g}(t) = \bar{S}_t \bar{f} = \lim_{\lambda\to\infty} S_t(\lambda, A, \bar{f})$$

with

$$S_t(\lambda, A, \bar{f}) = \lambda \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{(n-1)!} e^{n\lambda t} R(n\lambda, A) \bar{f}.$$

Now the A^k are bounded operators and hence the e^{tA_k} satisfy a condition of type (3.2) for every k; so we can apply the same representation to $\bar{g}^k(t) = e^{tA_k}\bar{f}^k$ to get

$$\bar{g}^k(t) = e^{tA_k}\bar{f}^k = \lim_{\lambda\to\infty} S_t(\lambda, A_k, \bar{f}).$$

Next we embed the finite-dimensional vector $\bar{g}^k(t)$, as any element of $\mathcal{H}_k$, in $\mathcal{H}$ simply by appending an infinite sequence of zeros and thus consider $\bar{g}^k(t)$ as an element of $\mathcal{H}$. Likewise, we can also consider $R(n\lambda, A_k)\bar{f}^k$ as an element of $\mathcal{H}$. Hence we can consider convergence of the $\bar{g}^k(t)$ as elements in $\mathcal{H}$.

Let λ_0 be large enough such that the resolvent $R(\lambda, A)$ is defined for all $\lambda > \lambda_0$. By Definition 3.5 and the assumed f-applicability,

$$\lim_{k\to\infty} R(n\lambda, A_k)\bar{f}^k = R(n\lambda, A)\bar{f}$$

for all $n \geq 1$ and $\lambda > \lambda_0$, where the limit is taken in $\mathcal{H}$. Having established this convergence, we invoke Lemma A.3 which states that then also

$$\lambda \sum_{n=1}^{\infty} (-1)^{n-1} \frac{e^{n\lambda t}}{(n-1)!} R(n\lambda, A_k) \bar{f}^k \longrightarrow \lambda \sum_{n=1}^{\infty} (-1)^{n-1} \frac{e^{n\lambda t}}{(n-1)!} R(n\lambda, A) \bar{f}.$$

Recall that our aim is to show that the $\bar{S}_t^k \bar{f}^k$ (considered as elements of $\mathcal{H}$) converge to $\bar{S}_t \bar{f}$, where $\bar{S}_t^k = e^{A_k t}$. Therefore, let $\varepsilon > 0$ and consider

$$\begin{aligned} \|\bar{S}_t^k \bar{f}^k - \bar{S}_t \bar{f}\| \leq \|\bar{S}_t^k \bar{f}^k - S_t(\lambda, A_k, \bar{f}^k)\| \\ + \|S_t(\lambda, A_k, \bar{f}^k) - S_t(\lambda, A, \bar{f})\| \\ + \|S_t(\lambda, A, \bar{f}) - \bar{S}_t \bar{f}\|. \end{aligned} \tag{3.7}$$

Since the S_t satisfy (3.2) and the S_t^k satisfy (3.3), we apply Corollary A.2 to both semigroups. So for all large λ, say $\lambda > \lambda_0$, the last term on the right-hand side in (3.7) is less than

$$D\|A\bar{f}\| e^{wt} \lambda^{-1/2} + \exp(-e^{\lambda t}) \|\bar{f}\|.$$

For the first term in (3.7), we have the upper bound

$$\|\bar{S}_t^k \bar{f}^k - S_t(\lambda, A_k, \bar{f}^k)\| \leq D_{\bar{f}^k} \|A_k \bar{f}^k\| e^{w_{\bar{f}^k} t} \lambda^{-1/2} + \exp(-e^{\lambda t}) \|\bar{f}^k\|.$$

Consider $\|A_k \bar{f}^k\|$. Since $A_k \bar{f}^k = P_k A P_k P_k \bar{f} = P_k A P_k \bar{f} = P_k A \bar{f}^k$, one finds $\|A_k \bar{f}^k\| \leq \|P_k\| \|A \bar{f}^k\| \leq \|A \bar{f}^k\|$. Note also that $\|\bar{f}^k\| \leq \|\bar{f}\|$. Hence the right-hand side of the last inequality has the first upper bound

$$D_{\bar{f}^k} \|A \bar{f}^k\| e^{w_{\bar{f}^k} t} \lambda^{-1/2} + \exp(-e^{\lambda t}) \|\bar{f}\|. \tag{3.8}$$

Under the assumptions (3.4) on w_f^u and (3.6) on α, we get for (3.8) an upper bound of the type encountered above, with $D = \sup_k D_{\bar{f}^k} < \infty$ under (3.4) as follows from the proof of (A.1). To be precise, (3.8) and thus also the first term in (3.7) is bounded by

$$D\alpha e^{w_f^u t} \lambda^{-1/2} + \exp(-e^{\lambda t}) \|\bar{f}\|.$$

It follows that the sum of the first and last term in (3.7) have an upper bound that tends to zero for $\lambda \to \infty$, uniformly in k. Choose then $\lambda > \lambda_0$ such that this sum is less than some $\varepsilon > 0$. For the chosen λ, the middle term in (3.7) can be made smaller than ε by choosing k larger than some $k_0 = k_0(\lambda_0)$ by Lemma A.3. Hence the total expression on the right of (3.7) is less than 2ε for all $k > k_0$. This concludes the proof. □

Theorem 3.7 dictates what steps should be followed to apply the approximation theory outlined in this section to a Feller process X on a state space D with generator $\mathcal{A}$. The first is to fix an appropriate basis $b(x)$ for the state space D of the process. Typically, Hermite (or other orthogonal) polynomials are instrumental here. The next step is to verify that the process is f-sequential for the $f \in \mathcal{P}$ that is of interest.

Obviously, bounded functions pose no difficulties. The examples in the remaining sections all deal with bounded functions f. Then one has to derive A column by column, based on $c^{(i)} = \mathcal{A}b_i$. Section 3.2 derives some useful results for Hermite polynomial bases. As a final step, one has to show that the applicability condition in Theorem 3.7 holds. We outline several strategies in the results below.

3.1 Results on FSM applicability to the inverse of resolvents

This section outlines some sufficient conditions on A and f that allow us to conclude that the FSM is f-applicable.

Applicability of the finite-section approximation has been extensively studied in the literature and sufficient conditions in a variety of settings have been established; we list a number of them. A well-known result is Polski's theorem, see e.g. Hagen et al. [15, Theorem 1.4], which states that applicability holds if the approximation $B_k P_k \to B$ (strongly) is *stable*, i.e., the norms of the inverses $\|B_k^{-1}\|$ are bounded, and if B is invertible. Refinements are given in Rabinovich et al. [25, Theorem 2.3] for banded and band-dominated operators, where stability is characterised by invertibility of some associated operators. Another important case where applicability holds (under conditions) is when dealing with infinite upper Hessenberg matrices; see for instance Robert and Santiago [26, Theorem 4]. A special case of this arises when dealing with (tri-diagonal) Jacobi matrices. More results can be found in Lindner [21, Sect. 5.2] where different (weaker) topologies are considered, like that induced by strict convergence.

Note that the stability condition on the norms of the resolvents $R(\lambda, A_k)$ is satisfied as soon as the A_k are such that their norms satisfy a bound as in (3.2), uniformly in k. Indeed, we then have

$$\|R(\lambda, A_k)\| \leq \int_0^\infty e^{-\lambda t} \|S_t^k\| \mathrm{d}t \leq \int_0^\infty e^{-\lambda t} C e^{wt} \mathrm{d}t = \frac{C}{\lambda - w}$$

for $\lambda > w$, independently of k.

The following result helps to establish strong convergence in specific cases that are of interest, for example an upper Hessenberg matrix (a matrix with all elements below its first sub-diagonal zero) with a bounded first sub-diagonal. This structure appears in the case of CIR bond pricing.

Proposition 3.8 *Consider an infinite-dimensional matrix B, also seen as an operator B from a domain $\mathcal{D}(B) \subseteq \ell^2$ to ℓ^2. If B is closed and, viewed as a matrix, has a strictly lower triangle which is banded with uniformly bounded elements, then $B_k P_k \to B$ strongly. Recall that a banded matrix B with bandwidth k has $B_{ij} = 0$ for $|i - j| > k$.*

Proof Our aim is to show $B_k P_k \to B$ strongly, i.e., the convergence in ℓ^2 of $B_k P_k x$ to Bx, for arbitrary $x \in \mathcal{D}(B)$. For the arguments that follow, we can without loss of generality assume that $\mathcal{D}(B) = \ell^2$, and so we assume that x is an arbitrary element of ℓ^2. Important is only that Bx is a well-defined element of ℓ^2.

Let $x \in \ell^2$ and $x^k = P_k x$; so $x = x^k + x^{\check{k}}$ with the projection error $x^{\check{k}}$. Consider

$$Bx - B_k P_k x = B(x^k + x^{\check{k}}) - B_k x^k = (B - B_k)x^k + Bx^{\check{k}}.$$

To show that $B_k P_k \to B$ strongly, we have to prove the convergence in ℓ^2 of the right-hand side of the above display. Consider first $Bx^{\check{k}}$. Suppose $Bx^{\check{k}} \to y$ (possibly along a subsequence) for $k \to \infty$. As $x^{\check{k}} \to 0$ in ℓ^2 for $k \to \infty$, the closedness of B implies $y = 0$.

It remains to check that $(B - B_k)x^k \to 0$ for $k \to \infty$. To that end, it is convenient to switch to coordinate representations of vectors and to a matrix representation of B. Let us write, with a little abuse of notation, x^k as

$$\begin{pmatrix} x^k \\ 0 \end{pmatrix},$$

where the x^k in the display is to be interpreted as a k-dimensional vector in $\mathbb{R}^k$. Put

$$B = \begin{pmatrix} B^{kk} & B^{k\check{k}} \\ B^{\check{k}k} & B^{\check{k}\check{k}} \end{pmatrix},$$

where $B^{kk} \in \mathbb{R}^{k\times k}$. Then we can represent $(B - B_k)x^k$ as

$$\left(\begin{pmatrix} B^{kk} & B^{k\check{k}} \\ B^{\check{k}k} & B^{\check{k}\check{k}} \end{pmatrix} - \begin{pmatrix} B^{kk} & 0 \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} x^k \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ B^{\check{k}k} x^k \end{pmatrix}.$$

So it remains to check that $B^{\check{k}k} x^k \to 0$ for $k \to \infty$. Suppose that the lower triangle of B is banded in the sense that $B_{ij} = 0$ for $i - j > d$ for some nonnegative integer d. Note that for $d = 0$, one finds that B is diagonal and hence the matrix $B^{\check{k}k}$ consists entirely of zeros, as this matrix has as first row the elements $(B_{k+1,1}, \ldots, B_{k+1,k})$, which are all zero when $d = 0$.

Therefore, we assume from here on $d > 0$ and additionally $k > d$ as $k \to \infty$. It follows that only the first d rows of $B^{\check{k}k}$ are possibly nonzero; all other rows are necessarily zero. The first d elements of $y_{k,d} := B^{\check{k}k} x^k$ are given by

$$B_{k,d} x_{k,d} := \begin{pmatrix} B_{k+1,k+1-d} & \cdots & \cdots & \cdots & B_{k+1,k} \\ 0 & B_{k+2,k+2-d} & \cdots & \cdots & B_{k+2,k} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & 0 & B_{k+d-1,k-1} & B_{k+d-1,k} \\ 0 & \cdots & \cdots & 0 & B_{k+d,k} \end{pmatrix} \begin{pmatrix} x_{k+1-d} \\ \vdots \\ x_k \end{pmatrix},$$

whereas the remaining elements are all zero. The matrices $B_{k,d}$ in the above display have the fixed size $d \times d$, and all their entries are bounded by a constant not depending on k by assumption. So the $B_{k,d}$ have uniformly bounded operator norms, and hence the ℓ^2-norm of $y_{k,d}$ is upper bounded by a constant times the ℓ^2-norm of $x_{k,d}$, which is $(\sum_{j=1}^{d} x_{k-d+j}^2)^{1/2}$ and smaller than the root of $\sum_{j=1}^{\infty} x_{k-d+j}^2$. Since $x \in \ell^2$, the latter quantity is the tail of a convergent series and hence vanishes for $k \to \infty$. □

The simplest case for applicability follows directly from Proposition 3.8 and occurs when we are dealing with a univariate polynomial process. For univariate polynomial processes, the corresponding matrix A is upper triangular.

Corollary 3.9 *Let A be as defined for Theorem 3.7. If A is upper triangular and $\bar{f}$ has a finite number of nonzero elements, then the FSM is f-applicable to the inverse $R(\lambda, A)^{-1} = \lambda I - A$.*

Proof We can take k and λ sufficiently large such that by assumption $\bar{f} = \bar{f}^k$ and that the inverse $R(\lambda, A_k)$ exists, since A_k is finite-dimensional.

Next, we write $\lambda I - A$ in block form as

$$\lambda I - A = \begin{pmatrix} \lambda I - A_k & -A' \\ 0 & \lambda I - A_\infty \end{pmatrix}.$$

Then by the properties of block inversion,

$$R(\lambda, A)\bar{f} = \begin{pmatrix} (\lambda I - A_k)^{-1} & (\lambda I - A_k)^{-1} A' (\lambda I - A_\infty)^{-1} \\ 0 & (\lambda I - A_\infty)^{-1} \end{pmatrix} \begin{pmatrix} \bar{f}^k \\ 0 \end{pmatrix} = R(\lambda, A_k)\bar{f}^k,$$

which concludes the proof. □

Beyond this simple case, we can show applicability for the Cox–Ingersoll–Ross bond price in Sect. 4.1. We defer the application of results to that section. For other examples, we show only numerical convergence.

3.2 Results on Hermite polynomials

As subsequent sections rely heavily on Hermite polynomials, we introduce some notation and properties. To denote the Hermite polynomial of order k, we use the notation $\mathrm{He}_k(x)$. We have (up to order four)

$$\begin{aligned} \mathrm{He}_0(x) &= 1, \\ \mathrm{He}_1(x) &= x, \\ \mathrm{He}_2(x) &= x^2 - 1, \\ \mathrm{He}_3(x) &= x^3 - 3x, \\ \mathrm{He}_4(x) &= x^4 - 6x^2 + 3. \end{aligned}$$

Per previous notation, these polynomials are extended to zero at Δ. Hermite polynomials are based on projection with a Gaussian weighting function with zero mean and unit variance. It is in our case more accurate to align the mean and variance with the unconditional mean and variance of the process. We allow two parameters, α for the variance (scaling) and β for the mean (location), and define

$$\mathrm{He}_k^{\alpha,\beta}(x) = \alpha^{\frac{n}{2}} \mathrm{He}_k\left(\frac{x - \beta}{\sqrt{\alpha}}\right).$$

The relationships in the following result are useful in applying an infinitesimal generator to Hermite polynomials of this kind.

Lemma 3.10 *Let $k, \ell \geq 0$ and $t > 0$. Then*

$$\begin{aligned}
\frac{d}{dx}\mathrm{He}_k^{\alpha,\beta}(x) &= k\mathrm{He}_{k-1}^{\alpha,\beta}(x),\\
\frac{d^\ell}{dx^\ell}\mathrm{He}_k^{\alpha,\beta}(x) &= \frac{k!}{(k-\ell)!}\mathrm{He}_{k-\ell}^{\alpha,\beta}(x),\\
x\mathrm{He}_k^{\alpha,\beta}(x) &= \alpha k\mathrm{He}_{k-1}^{\alpha,\beta}(x) + \beta\mathrm{He}_k^{\alpha,\beta}(x) + \mathrm{He}_{k+1}^{\alpha,\beta}(x),\\
e^x\mathrm{He}_k^{\alpha,\beta}(x) &= \sum_{\ell=0}^{\infty} e^{\beta+\frac{\alpha}{2}} \sum_{i=0}^{k\wedge\ell} \binom{k}{i} \frac{\alpha^{k-i}}{(\ell-i)!}\mathrm{He}_\ell^{\alpha,\beta}(x).
\end{aligned}$$

Proof See Appendix B. □

3.3 Numerical considerations

In all applications in the sections to follow, we use a standard implementation of the algorithm by Al-Mohy and Higham [2] to compute the action of the matrix exponential $e^{tA_k}\bar{f}^k$ directly for a grid of times, since it is much faster and numerically more stable than computing e^{tA_k} separately for several values of t before multiplying by $\bar{f}^k$. In cases where the interest is in multiple moments $\bar{f}^{k,1}, \ldots, \bar{f}^{k,m}$, we use the same algorithm on $e^{tA_k}\bar{F}^k$ with $\bar{F}^k := (\bar{f}^{k,1}, \ldots, \bar{f}^{k,m})$.

Table 1 compares the speed of this calculation for different matrix sizes and numbers of time steps. Even for very large matrices, the computation runs in subsecond time. Note that only one of these computations has to be executed for a given parametrisation. If the state x of the process changes, we only have to calculate $b^k(x)$

Table 1 Computation time (in milliseconds) of $e^{tA_k}F^k$ for various practically relevant sizes of A_k and F^k and a grid of times of size 1, 10 and 100, on four Intel(R) Xeon(R) CPU @ 2.20 GHz cores. Entries are generated randomly and scaled to prevent explosion. Numbers reported are medians of 10 runs

size A	size B	time steps		
		1	10	100
10×10	10×1	1.4	5.6	49.8
	10×10	1.0	6.0	56.0
20×20	20×1	1.0	5.3	51.6
	20×10	1.1	6.8	62.6
50×50	50×1	1.0	5.8	55.3
	50×10	1.3	8.5	77.1
100×100	100×1	2.4	13.3	121.6
	100×10	3.4	10.9	106.0
200×200	200×1	2.8	15.3	137.5
	200×10	4.3	28.1	271.9
500×500	500×1	4.7	24.5	173.7
	500×10	13.1	77.9	717.3

and get $b^k(x)^\top e^{tA_k}\bar{F}^k$ through multiplication (both negligible in speed compared to the matrix exponential). In the case of MLE, this means that the matrix exponential has to be computed only once per likelihood evaluation, even if we have a very long time series. In the case of scenario simulation, we can pre-compute the matrix exponential offline as part of the calibration.

In calibration and estimation exercises, the interest may be in the derivative of $\bar{g}(t)$ with respect to the parameters θ of the process which are encoded in A_k. That is, we are interested in $\frac{\partial}{\partial\theta_i}\bar{g}(t)$, with $\frac{\partial}{\partial\theta_i}A_k$ easy to derive from its definition. Using standard properties of the Fréchet derivative, we have

$$e^{\tilde{A}_k} = \begin{pmatrix} e^{A_k} & \frac{\partial}{\partial\theta_i}e^{A_k} \\ 0 & e^{A_k} \end{pmatrix}, \qquad \tilde{A}_k = \begin{pmatrix} A_k & \frac{\partial}{\partial\theta_i}A_k \\ 0 & A_k \end{pmatrix},$$

and therefore

$$e^{t\tilde{A}_k}\begin{pmatrix} 0 \\ \bar{f}^k \end{pmatrix} = \begin{pmatrix} e^{tA_k} & \frac{\partial}{\partial\theta_i}e^{tA_k} \\ 0 & e^{tA_k} \end{pmatrix}\begin{pmatrix} 0 \\ \bar{f}^k \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial\theta_i}\bar{g}(t) \\ \bar{g}(t) \end{pmatrix}.$$

Note how this computes both the value and the derivative in a single expanded matrix exponential.

4 Applications to short-rate models

The next two sections analyse several possible applications of polynomial moment approximation. We start with the bond price approximations of two popular short-rate models, before considering more complex credit-spread models in the next section.

Many popular short-rate models have the following structure: a Feller process X is specified as well as a function $r : D \to \mathbb{R}_+$ such that the short rate at time t is given by $r(X_t)$. In this context, the zero-coupon-bond price is given by the expectation

$$P(x,t,T) = \mathbb{E}\big[e^{-\int_t^T r(X_s)\mathrm{d}s}\big|X_t = x\big]. \tag{4.1}$$

As shown by Duffie et al. [14], this price is equivalent to the zeroth moment of a modified process with a generator $\mathcal{A}f(x) := \mathcal{A}_x f(x) - r(x)f(x)$, with $\mathcal{A}_x$ being the infinitesimal generator of the process X and $r(x)$ the killing rate at which the process jumps to the cemetery state Δ. Nonnegativity of r on D is required to ensure that the semigroup (S_t) is Feller, since Feller semigroups are contraction operators. When there is a constant lower bound, i.e., when the discount rate can be written as $r(x) = \underline{r} + r'(x)$ with r' nonnegative on D, then we can bring $e^{-\underline{r}(T-t)}$ outside the expectation in (4.1). However, we have seen that the approximation may also work for negative discounting directly. With the semigroup (S_t) corresponding to $\mathcal{A}$, the bond price with respect to the basis $b(x) = (\mathrm{He}_0^{\alpha,\beta}(x), \mathrm{He}_1^{\alpha,\beta}(x), \ldots)^\top$ is given by

$$P(x,t,T) = S_{T-t}f(x), \qquad \bar{f} = e_1 = (1,0,\ldots)^\top.$$

Both the Cox–Ingersoll–Ross (CIR) and the Black–Karasinski model fall in this class. In the CIR case, a closed-form solution exists, making it an excellent reference

case for testing the approximation. In the Black–Karasinski case, we compare with numerical methods since an analytical solution does not exist.

4.1 The Cox–Ingersoll–Ross bond price

The CIR one factor short-rate model in [12] is a popular model to price interest-rate derivatives. Its state space is the positive real line $D = \mathbb{R}_+$ so that negative rates are avoided. A closed-form solution exists for the price of a (zero-coupon) bond to benchmark our approximation.

The CIR short-rate dynamics follow the SDE

$$\mathrm{d}X_t = \theta(\mu - X_t)\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t, \qquad r(X_t) = X_t. \tag{4.2}$$

In (4.2), the initial state X_0 is positive, and so are the parameters μ, σ and θ. Furthermore, $2\theta\mu > \sigma^2$ to ensure that the process remains positive. The infinitesimal generator of the modified process is

$$\mathcal{A}f(x) = \theta(\mu - x)\frac{\partial f}{\partial x}(x) + \frac{1}{2}\sigma^2 x\frac{\partial^2 f}{\partial x^2}(x) - xf(x),$$

where the last term is the adjustment that allows us to compute the bond price as the first moment of the process. Applying this infinitesimal generator to the basis elements $b_{i+1}(x) = \mathrm{He}_i^{\alpha,\beta}(x)$, $0 \le i \le k-1$, gives

$$\begin{aligned}
\mathcal{A}\mathrm{He}_i^{\alpha,\beta}(x) = &-\mathrm{He}_{i+1}^{\alpha,\beta}(x)\\
&-(\theta i + \beta)\mathrm{He}_i^{\alpha,\beta}(x)\\
&+\left(\frac{1}{2}\sigma^2 i(i-1) + (\theta\mu - \theta\beta - \alpha)i\right)\mathrm{He}_{i-1}^{\alpha,\beta}(x)\\
&+\left(\frac{1}{2}\sigma^2\beta - \theta\alpha\right)i(i-1)\mathrm{He}_{i-2}^{\alpha,\beta}(x)\\
&+\frac{1}{2}\sigma^2\alpha i(i-1)(i-2)\mathrm{He}_{i-3}^{\alpha,\beta}(x).
\end{aligned}$$

From this expression, we can columnwise identify the matrix A and then A_k by considering the $k \times k$ finite section. For example, for $k = 4$, we have

$$A_4 = \begin{pmatrix}
-\beta & \theta\mu - \theta\beta - \alpha & \beta\sigma^2 - 2\theta\alpha & 3\sigma^2\alpha\\
-1 & -(\theta + \beta) & \sigma^2 + 2(\theta\mu - \theta\beta - \alpha) & 3\beta\sigma^2 - 6\theta\alpha\\
0 & -1 & -2\theta - \beta & 3\sigma^2 + 3(\theta\mu - \theta\beta - \alpha)\\
0 & 0 & -1 & -3\theta - \beta
\end{pmatrix}.$$

Choosing $\alpha = \frac{\mu\sigma^2}{2\theta}$ and $\beta = \mu$ gives the alignment of the Hermite weighting function with the unconditional mean and variance of the process.

First, the (killed) process X is f-sequential because the contingent claim $f \equiv 1$ in this context is bounded. Second, we note that $\bar{f} = e_1$ so that (3.5) is automatically

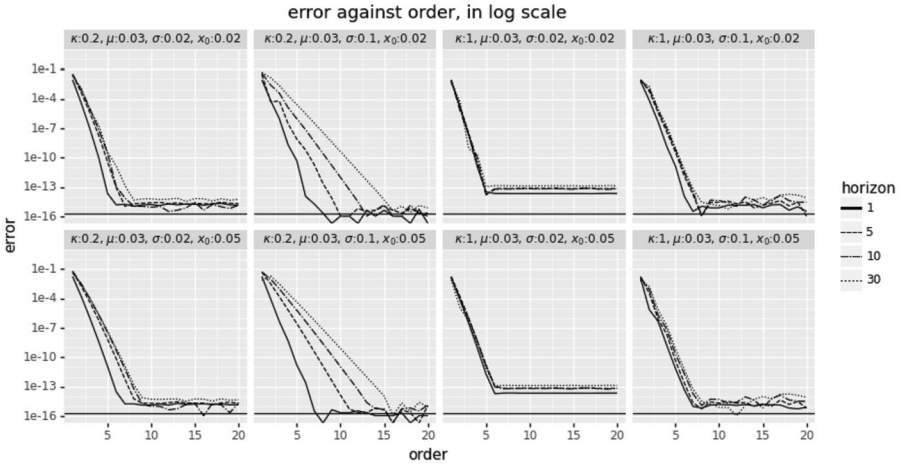

Fig. 2 Absolute approximation error of zero-coupon-bond price yield. Different lines represent tenors. The horizontal lines indicate machine precision

satisfied and thus also (3.6). We assume that (3.3) holds, but were not able to prove this formally. In numerical experiments, we were not able to find a violation. Second, A exists and is upper Hessenberg with bounded lower diagonal; so it has a banded and bounded strict lower triangle and Proposition 3.8 can be applied to show strong convergence. Next, $\lambda I - A_k$ is invertible for all $k > k_0$ for some k_0 (see the proof of Corollary 3.9). For Polski's theorem, we also require that $\|(\lambda I - A_k)^{-1}\|$ are bounded uniformly in k, except for banded matrices where this requirement can be dropped; see Rabinovich et al. [25, Theorem 2.3]. Therefore the f-applicability condition in Theorem 3.7 is also satisfied.

In all subsequent examples that we consider, the f-applicability of the FSM to $\lambda I - A$ could not be established with the tools developed in Sect. 3. Instead, we show convergence numerically.

To show the accuracy of the proposed approximation, we compare it with the analytical solution for eight sets of different but typical parameters. Figure 2 shows that the error decreases exponentially as the approximation order k increases. For most sets of parameters, it converges to one or two orders of magnitude above machine precision, and remains stable as the order increases.

4.2 The Black–Karasinski bond price

The Black–Karasinski short-rate model is similar in structure to the CIR model, but assumes that short rates follow an exponential OU process. No analytical solution is available; hence finding an efficient and accurate approximation has received considerable academic attention. This section shows that our approximation is accurate against a PDE benchmark.

Black and Karasinski [7] assume the short rate (r_t) has dynamics

$$\mathrm{d}\ln r_t = \theta(\mu - \ln r_t)\mathrm{d}t + \sigma \mathrm{d}W_t.$$

In our setup, this translates to

$$\mathrm{d}X_t = \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad r(X_t) = e^{X_t}.$$

The process has unconditional mean zero to align with the Hermite polynomial approximation density. The infinitesimal generator of the modified process is

$$\mathcal{A}f(x) = \theta(\mu - x)\frac{\partial f}{\partial x}(x) + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2}(x) - e^x f(x).$$

Applying this infinitesimal generator to the basis elements $b_{i+1}(x) = \mathrm{He}_i^{\alpha,\beta}(x)$ with $0 \le i \le k-1$ gives

$$\begin{aligned}\mathcal{A}\mathrm{He}_i^{\alpha,\beta}(x) = &-\theta i \mathrm{He}_i^{\alpha,\beta}(x) \\ &+\theta i(\mu - \beta)\mathrm{He}_{i-1}^{\alpha,\beta}(x) \\ &+\left(\frac{1}{2}\sigma^2 + \theta\alpha\right) i(i-1)\mathrm{He}_{i-2}^{\alpha,\beta}(x) \\ &+\sum_{j=0}^{\infty} e^{\beta + \frac{\alpha}{2}} \sum_{\ell=0}^{i\wedge j} \binom{j}{\ell} \frac{\alpha^{j-\ell}}{(i-\ell)!} \mathrm{He}_j^{\alpha,\beta}(x).\end{aligned}$$

From this expression, we can again columnwise identify the matrix generator A and its $k \times k$ finite sections A_k. We may assume that $\alpha = \frac{\sigma^2}{2\theta}$ and $\beta = \mu$, in line with the unconditional mean and variance of the OU process.

We calculate a set of zero-coupon-bond yields with different parameters and use a PDE solver to benchmark the approximation quality. The unconditional distribution

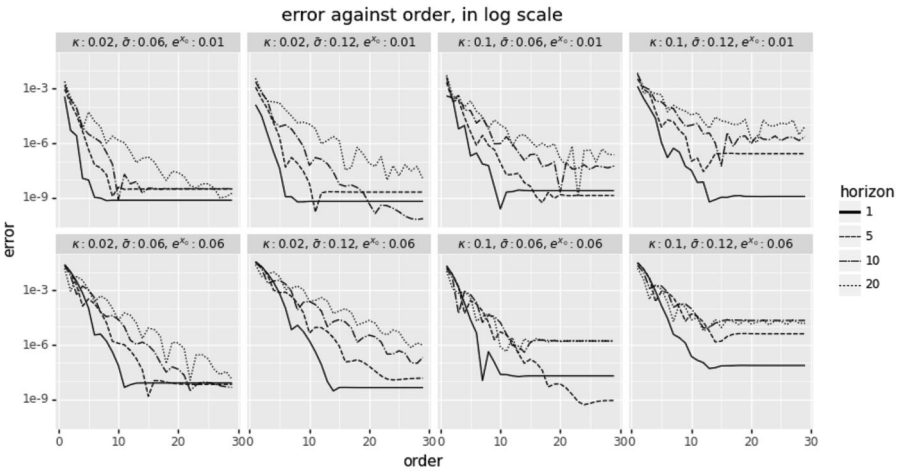

Fig. 3 Absolute approximation error of zero-coupon-bond price yield against the PDE solution. Different lines represent tenors. For different parameters, the unconditional mean $\bar{\mu}$ is kept constant at 0.03

of X_t is Gaussian, with unconditional mean μ and variance $\frac{\sigma^2}{2\theta}$. Thus the steady-state distribution of the short rate $r(X_t) = e^{X_t}$ has mean $\bar{\mu} = \exp(\mu + \frac{\sigma^2}{4\theta})$ and variance $\bar{\sigma}^2 = (\exp(\frac{\sigma^2}{2\theta}) - 1)\exp(2\mu + \frac{\sigma^2}{2\theta})$. We fix $\bar{\mu} = 0.03$ and vary the other parameters to obtain realistic alternative sets of parameters:

1) values of $x = \ln r_0$: $\ln 0.01$, $\ln 0.03$ and $\ln 0.06$;
2) values of θ: 0.02 and 0.1;
3) values of $\bar{\sigma}$: 6% and 12%.

Figure 3 outlines the approximation error for maturities 1, 5, 10 and 20 years, as a function of the approximation order. The approximation error versus the PDE solution decreases with the order. For order $k = 20$, the error is no greater than 1 bps. We do note that we cannot separate the accuracy of the PDE solver from that of the polynomial approximation. The accuracy of the polynomial approximation may very well be close to machine precision for higher orders.

5 Applications to credit derivatives

We follow the setup of the generalised Markovian model of credit rating migrations introduced by Lando [19]. This model assumes that companies migrate independently within a set of m ratings, e.g. {AAA, AA, ..., CCC, D}, where AAA is the highest quality rating and D represents default, or {IG, HY, D} for investment grade and high-yield bonds. A company's rating follows a continuous-time Markov chain (R_t) with the ratings as states, and with an $m \times m$ generator matrix $Q(Y_t)$ that depends on a latent driving process (Y_t) of state variables.

Consider the $m \times m$ rating migration probability matrix conditionally on the full history of state variables, i.e.,

$$P_{ij}^Y(t) := \mathbb{P}[R_t = j | R_0 = i, \mathcal{F}_t^Y],$$

where $\mathcal{F}_t^Y := \sigma(Y_s, 0 \le s \le t)$, $t \ge 0$, is the natural filtration of the stochastic process (Y_t). Then $P^Y(t)$ satisfies the Kolmogorov forward equation

$$\partial_t P^Y(t) = P^Y(t) Q(Y_t), \qquad P^Y(0) = I_m.$$

In order to derive credit spreads and rating migration probabilities, the interest is in the *rating migration matrix*

$$P_{ij}(t, y) := \mathbb{P}_y[R_t = j | R_0 = i] = \mathbb{E}_y[P_{ij}^Y(t)].$$

As shown by Lando [19], in the specific case that the $Q(Y_t)$ commute and are diagonalisable, we can solve the Kolmogorov forward equation and write

$$P^Y(t) = e^{\int_0^t Q(Y_s)\mathrm{d}s} = B e^{\int_0^t D(Y_s)\mathrm{d}s} B^{-1} = B \mathrm{diag}_i\left(e^{\int_0^t D_{ii}(Y_s)\mathrm{d}s}\right) B^{-1},$$

with diagonalisation $Q(y) = BD(y)B^{-1}$, where $D(y)$ is a diagonal matrix of (non-positive) eigenvalues. This strategy uses the fact that commuting diagonalisable matrices are simultaneously diagonalisable, i.e., share the matrix B. Taking the expectation results in a set of bond-price-like formulas that can be solved analytically in certain cases, i.e.,

$$P(t, y) = \mathbb{E}_y\big[e^{\int_0^t Q(Y_s)\mathrm{d}s}\big] = B\mathrm{diag}_i\big(\mathbb{E}_y\big[e^{\int_0^t D_{ii}(Y_s)\mathrm{d}s}\big]\big)B^{-1}. \tag{5.1}$$

In this section, we do *not* assume a commuting property of the generators and use the proposed approximation strategy to calculate the rating migration matrix. To this end, we define the basis vector-valued process Z with $Z_t = e_{R_t}$ with the state space $E = \{e_1, \ldots, e_m\}$ of m-dimensional basis vectors. We assume that (Y_t) follows an n-dimensional time-homogeneous Itô diffusion with state space D'. The SDE for the joint process $X := (Y, Z)$ is

$$\begin{aligned}
\mathrm{d}Y_t &= \mu(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t, \\
\mathrm{d}Z_t &= \sum_{i=1}^m Z_{i,t-} \sum_{j\neq i} (e_j - e_i)\mathrm{d}N_t^{ij},
\end{aligned}$$

where the (N_t^{ij}) are Poisson processes with intensity $\mathbb{E}[\mathrm{d}N_t^{ij}|\mathcal{F}_t] = Q(Y_t)\mathrm{d}t$, where the filtration $(\mathcal{F}_t)$ is generated by the processes (R_t) and (Y_t). Intuitively, if the Markov chain (R_t) is in state i at time t, then $Z_{i,t-} = 1$ and it migrates to state $j \neq i$ with intensity $Q_{ij}(Y_t)$. A jump to state j modifies Z_t by subtracting the current state e_i and adding the new state e_j. It follows from basic manipulation that

$$\mathrm{d}Z_t = Q(Y_t)^\top Z_{t-}\mathrm{d}t + \mathrm{d}M_t$$

with a martingale (M_t) for the filtration $(\mathcal{F}_t)$. To see this, use $Q(Y_t)1 = 0$ to get

$$\begin{aligned}
\mathbb{E}[\mathrm{d}Z_t|\mathcal{F}_t] &= \sum_{i=1}^m Z_{i,t-} \sum_{j\neq i} (e_j - e_i)\mathbb{E}[\mathrm{d}N_t^{ij}|\mathcal{F}_t] \\
&= \sum_{i=1}^m Z_{i,t-} \sum_{j=1}^m (e_j - e_i)Q_{ij}(Y_t)\mathrm{d}t \\
&= \sum_{i=1}^m Z_{i,t-} \sum_{j=1}^m e_j Q_{ij}(Y_t)\mathrm{d}t - \sum_{i=1}^m Z_{i,t-} \sum_{j=1}^m e_i Q_{ij}(Y_t)\mathrm{d}t \\
&= \sum_{j=1}^m \Big(\sum_{i=1}^m Z_{i,t-}Q_{ij}(Y_t)\Big)e_j\,\mathrm{d}t - \sum_{i=1}^m Z_{i,t-}e_i\big(Q(Y_t)1\big)_i\,\mathrm{d}t \\
&= \sum_{j=1}^m \big(Z_{t-}^\top Q(Y_t)e_j\big)e_j\,\mathrm{d}t \\
&= \sum_{j=1}^m \big(Q(Y_t)^\top Z_{t-}\big)_j e_j\,\mathrm{d}t \\
&= Q(Y_t)^\top Z_{t-}\mathrm{d}t.
\end{aligned}$$

In this setting, we can express the rating migration matrix as an expectation that conforms to our approximation approach because

$$P_{ij}(t, y) = \mathbb{P}_y[Z_t = e_j | Z_0 = e_i] = \mathbb{E}_{y,e_i}[\langle e_j, Z_t\rangle] = S_t f(y, e_i),$$

where $f(y, z) = z_j$. In order to apply the approximation, we need the generator of the process X. If Y has generator $\mathcal{A}^y$, then by standard arguments, the generator of the process X is

$$\mathcal{A}f(y, z) = \mathcal{A}^y f(y, z) + \mathcal{A}^z f(y, z), \qquad \mathcal{A}^z f(y, z) := z^\top Q(y) \begin{pmatrix} f(y, e_1) \\ \vdots \\ f(y, e_m) \end{pmatrix}.$$

It is easy to see that with $b(y)$ an appropriate basis for Y, $b(x) := b(y)\otimes z$ is appropriate for X. In what follows, we assume $b(y) = (\mathrm{He}_0^{\alpha,\beta}(y), \mathrm{He}_1^{\alpha,\beta}(y), \mathrm{He}_2^{\alpha,\beta}(y), \ldots)$ so that $b(x) = (\mathrm{He}_0^{\alpha,\beta}(y)\otimes z, \mathrm{He}_1^{\alpha,\beta}(y)\otimes z, \mathrm{He}_2^{\alpha,\beta}(y)\otimes z, \ldots)$. With ℓ being the highest power of y included, this basis has for $n = 1$ dimension ℓm, for $n = 2$ dimension $\ell(\ell + 1)m/2$, and for general $n > 1$, the dimension is $\sum_{i=0}^{\ell-1} \binom{n+i-1}{n-1} m$.

The steps in approximating $P(t, y)$ are based on choosing moments. With the basis in place, we can derive A and A_k in the usual way. Note that $P_{ij}(t, y) = S_t f(y, e_i)$ with $f(y, z) = z_j = \langle (e_j, 0, \ldots), b(y, z)\rangle$. Again f is bounded so that the process is f-sequential. Therefore, we can approximate

$$P_{ij}(t, y) \approx \langle e^{tA_k}(e_j, 0)^\top, b^k(y, e_i)\rangle,$$
$$P(t, y) \approx \left(b^{k/m}(y)^\top \otimes I_m\right) e^{tA_k}(I_m, 0)^\top.$$

5.1 Migrations driven by multivariate CIR processes

Consider the process $X = (Y, Z)$ specified on the domain $D = \mathbb{R}_+^n \times E$, where $E = \{e_1, \ldots, e_m\}$ so that $d = n + m$, through its SDE

$$\mathrm{d}Y_t = K(\mu - Y_t)\mathrm{d}t + \mathrm{diag}_i(\sigma_i \sqrt{Y_{i,t}})\mathrm{d}W_t,$$
$$\mathrm{d}Z_t = Q(Y_t)^\top Z_t \mathrm{d}t + \mathrm{d}M_t,$$

where (Y_t) follows a multivariate CIR process with $n \times n$ mean reversion speed K to means μ. In case K is diagonal, the $(Y_{i,t})$ are n independent CIR processes. We choose the generator matrix $Q(y) := \sum_{i=1}^n y_i Q_i$, where the Q_i are generator matrices of continuous-time Markov chains. Since (Y_t) is nonnegative, $Q(Y_t)$ is a well-defined generator matrix.

In this specific case, the generator of the process X is

$$\mathcal{A}f(y, z) = \sum_{i=1}^n \left(\frac{1}{2}\sigma_i^2 y_i \frac{\partial^2 f(y, z)}{\partial y_i^2} + e_i^\top K(\mu - y)\frac{\partial f(y, z)}{\partial y_i} \right)$$
$$+ \sum_{i=1}^n y_i z^\top Q_i \begin{pmatrix} f(y, e_1) \\ \vdots \\ f(y, e_m) \end{pmatrix}$$

Applying $\mathcal{A}^y$ to elements of $b^k(y)$ always returns an order of at most k; hence A_y^1 is (block) upper triangular. However, A^z is not upper triangular.

5.1.1 Migrations driven by a univariate CIR process

Arvanitis et al. [5] apply the model above to the univariate case $n = 1$. With the diagonalisation $Q_1 = BDB^{-1}$, we use (5.1) to obtain

$$P(t, y) = B\mathrm{diag}_i\left(\mathbb{E}_y\left[e^{\int_0^t D_{ii}Y_s\mathrm{d}s}\right]\right)B^{-1}.$$

Since Y follows a CIR process and D_{ii} are the nonpositive eigenvalues of Q_1, the term $-YD_{ii}$ is either 0 or follows a CIR process. Thus each diagonal element is either 1 or a CIR bond price, which has an analytical solution. Then $P(t, y)$ has an analytical solution that we can use as a benchmark for our approximation.

To test the accuracy of the proposed approximation, we choose the same Markov chain generator matrix Q_1 as Jarrow et al. [18, Example 1], namely

$$Q_1 = \begin{pmatrix} -0.11 & 0.1 & 0.01 \\ 0.05 & -0.15 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Figure 4 plots the mean absolute approximation error across all matrix entries against the order $\ell = k/m$ of the approximation, for various parameters of the CIR process and maturities. We can see that the error is decreasing exponentially as the order increases. The convergence appears to be faster for shorter horizons. All errors converge to one or two orders above machine precision.

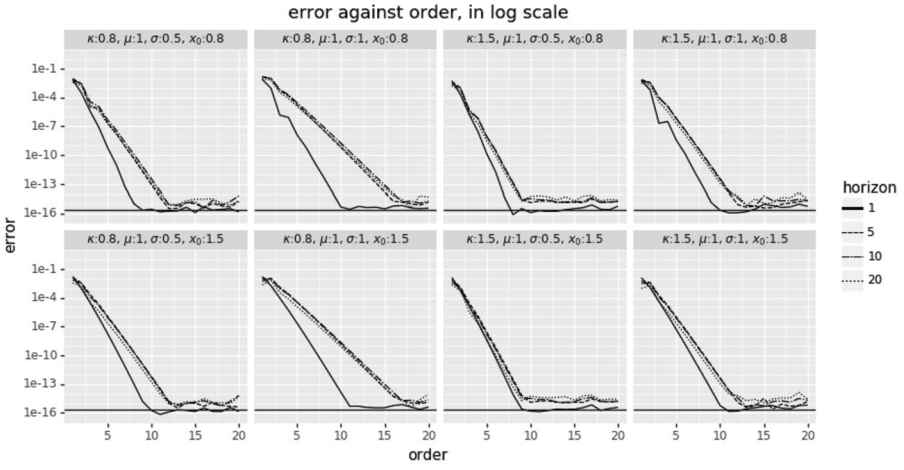

Fig. 4 Dimension one, migration probability elementwise mean absolute error against different orders, compared with analytical solution, in log scale. Different panels stand for different parameters. Different lines stand for different time horizons. Horizontal lines stand for machine precision

5.1.2 Migrations driven by a bivariate CIR process, commuting case

Hurd and Kuznetsov [16] apply the same credit model to the bivariate case $n = 2$. In order to ensure tractability, they specify Q_2 to reflect an additional default migration that is the same for all ratings. This second generator matrix can also be interpreted as a liquidity premium. The specific structure of this second matrix ensures that Q_1 and Q_2 commute. With the additional assumption that K is diagonal, i.e., the two driving CIR processes are independent, we can write

$$P(t, y) = \mathbb{E}_y\left[e^{\int_0^t (Y_{1,s}Q_1 + Y_{2,s}Q_2)\mathrm{d}s}\right] = \mathbb{E}_{y_1}\left[e^{\int_0^t Y_{1,s}Q_1\mathrm{d}s}\right]\mathbb{E}_{y_2}\left[e^{\int_0^t Y_{2,s}Q_2\mathrm{d}s}\right]$$

and apply the univariate pricing strategies.

We keep Q_1 the same as in the previous example and follow [16, Sect. 7] to define Q_2, i.e.,

$$Q_1 = \begin{pmatrix} -0.11 & 0.1 & 0.01 \\ 0.05 & -0.15 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Q_2 = \begin{pmatrix} -0.01 & 0 & 0.01 \\ 0 & -0.01 & 0.01 \\ 0 & 0 & 0 \end{pmatrix}.$$

We define the other parameters as $K_{\text{big}} = \text{diag}(1.5, 1.5)$, $K_{\text{small}} = \text{diag}(0.8, 0.8)$, $\sigma_{\text{big}} = (1.0, 1.0)$, $\sigma_{\text{small}} = (0.5, 0.5)$, $x_{\text{big}} = (1.2, 1.2)^\top$ and $x_{\text{small}} = (0.8, 0.8)^\top$. The parenthesis lists the parameters of the two independent CIR processes. Figure 5 shows the same pattern of approximation quality as in the univariate case.

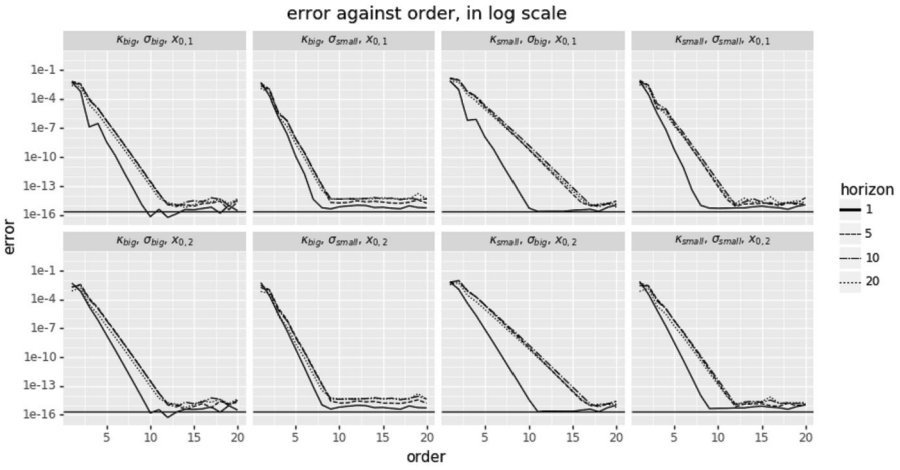

Fig. 5 Dimension two when analytical solution exists, migration probability elementwise mean absolute error against different orders, compared with analytical solution, in log scale. Different panels stand for different parameters. Different lines stand for different time horizons. Horizontal lines stand for machine precision

5.1.3 Migrations driven by a bivariate CIR process, non-commuting case

We now consider a case for which no analytical solution exists, namely when both the independence and commutativity requirements fail. Let Q_1 and Q_2 be upper and lower triangular, respectively. This means that Y_1 and Y_2 represent the scaling processes that accelerate the speed of upgrades and downgrades separately. This model can capture the important stylised fact that with the business cycle, upgrades tend to slow down when downgrades speed up, and vice versa. The details of this model are discussed in an accompanying empirical paper [6].

Since no analytical solution exists, we benchmark the approximation against a binomial tree. We choose Q_1 and Q_2 to be an upper and lower triangular decomposition of the transition matrix in Jarrow et al. [18, Example 1], i.e.,

$$Q_1 = \begin{pmatrix} -0.11 & 0.1 & 0.01 \\ 0 & -0.1 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Q_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0.05 & -0.05 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

These matrices do not commute. In fact, using Böttcher and Wenzel[9, Conjecture 1.2], we get the Frobenius norm inequality $\|Q_1Q_2 - Q_2Q_1\|_F \leq \sqrt{2}\|Q_1\|_F\|Q_2\|_F$. This inequality gives rise to a measure of non-commutativity with values in [0, 1] for non-trivial matrices, namely

$$\frac{\|Q_1Q_2 - Q_2Q_1\|_F}{\sqrt{2}\|Q_1\|_F\|Q_2\|_F} = 0.48.$$

This value allows us to conclude that these matrices are strongly non-commuting.

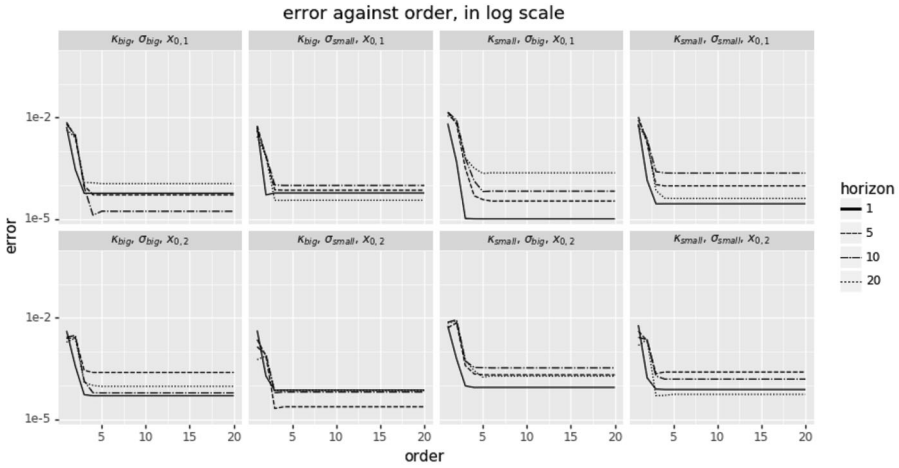

Fig. 6 Dimension two when analytical solution does not exist, migration probability elementwise mean absolute error against different orders, compared with MC simulation, in log scale. Different panels stand for different parameters. Different lines stand for different time horizons

The other parameters are given by $K_{\text{big}} = \begin{pmatrix} 1.5 & 0.4 \\ 0.4 & 1.5 \end{pmatrix}$, $K_{\text{small}} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$, $\sigma_{\text{big}} = (1.0, 1.0)$, $\sigma_{\text{small}} = (0.5, 0.5)$, $x_{\text{big}} = (1.2, 1.2)^\top$ and $x_{\text{small}} = (0.8, 0.8)^\top$. Note that the matrix K is no longer diagonal and can induce correlation between the upgrade and downgrade speed processes.

Figure 6 shows the mean absolute approximation error against the approximation order $\ell = k/m$. It is worth noting that the errors converge to bounds within the margin of error expected from the Monte Carlo simulation. Overall, we see the same pattern of exponential decline in the error, although as expected, a Monte Carlo-induced lower bound is hit sooner than when comparing against an analytical benchmark.

Appendix A: Supporting lemmas

The proof of Theorem 3.7 relies on two lemmas that are presented below.

The first lemma combines an inversion result à la Phragmén–Doetsch for Laplace transforms (see Arendt et al. [4, Theorem 2.3.2]) with Phragmén's approximation result for semigroups of operators (see Neubrander [23]). The general setting is that (S_t) is a strongly continuous semigroup acting on a Banach space $\mathcal{P}$ with generator $\mathcal{A}$ having domain $\mathcal{D}(\mathcal{A})$. Let $R(\lambda, \mathcal{A})$ denote the resolvent, which exists for all sufficiently large λ. Recall that (3.2) says that there exist $C \geq 1$ and $w \geq 0$ with $\|S_t\| \leq Ce^{wt}$.

In the proof of Lemma A.1 below, we need the quantity

$$\tilde{S}_t(\lambda, \mathcal{A}, f) := \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n!} e^{n\lambda t} R(n\lambda, \mathcal{A}) f.$$

A first result, adapted from Neubrander [23] and Arendt et al. [4, Theorem 2.3.2], is that for $f \in \mathcal{P}$ and $\lambda > w$, one has

$$\left\| \int_0^t S_u f \, \mathrm{d}u - \tilde{S}_t(\lambda, \mathcal{A}, f) \right\| \leq D \|f\| e^{wt} \lambda^{-1/2}, \tag{A.1}$$

where D is a constant possibly depending on t, and w is as in (3.2). To see that (A.1) is valid, we follow [23] with slightly different arguments. First we recall that [23, p. 106] shows that with $e_\lambda(x) = \exp(-e^{\lambda x})$ for $x > 0$, one has

$$\tilde{S}_t(\lambda, \mathcal{A}, f) = \int_0^\infty \big(1 - e_\lambda(t-u)\big) S_u f \, \mathrm{d}u.$$

Note the inequalities $0 < e_\lambda(x) < 1$ and $0 < 1 - e_\lambda(x) < 1$ and $1 - e_\lambda(x) < e^{\lambda x}$. They are used in the chain of norm inequalities below. But first we consider

$$\begin{aligned} \tilde{S}_t(\lambda, \mathcal{A}, f) - \int_0^t S_u f \, \mathrm{d}u &= \int_0^\infty \big(1 - e_\lambda(t-u)\big) S_u f \, \mathrm{d}u - \int_0^t S_u f \, \mathrm{d}u \\ &= -\int_0^{t-\delta} e_\lambda(t-u) S_u f \, \mathrm{d}u - \int_{t-\delta}^t e_\lambda(t-u) S_u f \, \mathrm{d}u \\ &\quad + \int_t^\infty \big(1 - e_\lambda(t-u)\big) S_u f \, \mathrm{d}u. \end{aligned}$$

Taking norms, we deduce for $0 < \delta < t$ and $\lambda > w$ that

$$\begin{aligned}
\left\| \tilde{S}_t(\lambda, \mathcal{A}, f) - \int_0^t S_u f \, \mathrm{d}u \right\| &\leq \int_0^{t-\delta} e_\lambda(t-u) \| S_u \| \| f \| \mathrm{d}u \\
&\quad + \int_{t-\delta}^{t} e_\lambda(t-u) \| S_u \| \| f \| \mathrm{d}u \\
&\quad + \int_t^\infty \big(1 - e_\lambda(t-u)\big) \| S_u \| \| f \| \mathrm{d}u \\
&\leq C e_\lambda(\delta) \int_0^{t-\delta} e^{wu} \| f \| \mathrm{d}u + C \int_{t-\delta}^{t} e^{wu} \| f \| \mathrm{d}u \\
&\quad + C \int_t^\infty e^{\lambda(t-u)} e^{wu} \| f \| \mathrm{d}u \\
&\leq C \| f \| \left(e_\lambda(\delta) t e^{wt} + \delta e^{wt} + e^{wt} \frac{1}{\lambda - w} \right) \\
&= C \| f \| e^{wt} \left(e_\lambda(\delta) t + \delta + \frac{1}{\lambda - w} \right).
\end{aligned}$$

Next we take $\delta \downarrow 0$ such that $\delta\lambda \to \infty$. Specifically, we choose $\delta = \lambda^{-1/2}$. As then $e_\lambda(\delta)$ and $\frac{1}{\lambda - w}$ tend to zero much faster than δ, we can bound the last term in the above display by $D \| f \| e^{wt} \lambda^{-1/2}$ for some constant D (which may depend on t and w).

Lemma A.1 *Recall* (3.2). *Let* $t > 0$ *and*

$$S_t(\lambda, \mathcal{A}, f) := \lambda \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{(n-1)!} e^{n\lambda t} R(n\lambda, \mathcal{A}) f. \tag{A.2}$$

Then for some $w \geq 0$ *and* $D \geq 1$, *we have for all* λ *big enough and all* $f \in \mathcal{D}(\mathcal{A})$ *that*

$$\| S_t f - S_t(\lambda, \mathcal{A}, f) \| \leq D \| \mathcal{A} f \| e^{wt} \lambda^{-1/2} + \exp(-e^{\lambda t}) \| f \|.$$

Consequently, $\| S_t f - S_t(\lambda, \mathcal{A}, f) \| \to 0$ *for* $\lambda \to \infty$.

Proof First we note that the sum in (A.2) is finite. This follows from Neubrander [23] and Arendt et al. [4, Theorem 2.3.2]. Following the line of argument in [23], we have for $f \in \mathcal{D}(\mathcal{A})$ the equality $R(n\lambda, \mathcal{A}) \mathcal{A} f = n\lambda R(n\lambda, \mathcal{A}) f - f$ and hence

$$\begin{aligned}
\tilde{S}_t(\lambda, \mathcal{A}, \mathcal{A} f) &= S_t(\lambda, \mathcal{A}, f) - \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n!} e^{n\lambda t} f \\
&= S_t(\lambda, \mathcal{A}, f) + \big(\exp(-e^{\lambda t}) - 1 \big) f.
\end{aligned}$$

Using $S_t f = f + \int_0^t S_u \mathcal{A} f \, \mathrm{d}u$ for $f \in \mathcal{D}(\mathcal{A})$, we develop, using (A.1) in the last inequality,

$$\begin{aligned}
&\|S_t f - S_t(\lambda, \mathcal{A}, f)\| \\
&= \left\| f + \int_0^t S_u \mathcal{A} f \, \mathrm{d}u - \tilde{S}_t(\lambda, \mathcal{A}, \mathcal{A} f) + \tilde{S}_t(\lambda, \mathcal{A}, \mathcal{A} f) - S_t(\lambda, \mathcal{A}, f) \right\| \\
&= \left\| f + \int_0^t S_u \mathcal{A} f \, \mathrm{d}u - \tilde{S}_t(\lambda, \mathcal{A}, \mathcal{A} f) + \left(\exp(-e^{\lambda t}) - 1 \right) f \right\| \\
&= \left\| \int_0^t S_u \mathcal{A} f \, \mathrm{d}u - \tilde{S}_t(\lambda, \mathcal{A}, \mathcal{A} f) + \exp(-e^{\lambda t}) f \right\| \\
&\leq \left\| \int_0^t S_u \mathcal{A} f \, \mathrm{d}u - \tilde{S}_t(\lambda, \mathcal{A}, \mathcal{A} f) \right\| + \exp(-e^{\lambda t}) \| f \| \\
&\leq D \| \mathcal{A} f \| e^{wt} \lambda^{-1/2} + \exp(-e^{\lambda t}) \| f \|.
\end{aligned}$$

□

If the space $\mathcal{P}$ is the Hilbert space as in Sects. 2 and 3, we can write the counterpart of Lemma A.1 for sequences in $\mathcal{H}$. If the S_t form the strongly continuous transition semigroup on $\mathcal{P}$ of a Feller process, they are all expectations and we can take $C = 1$ and $w = 0$ in (3.2). The same is true for the induced semigroup of operators $\bar{S}_t$ on $\mathcal{H}$. This leads to the following variation on Lemma A.1.

Corollary A.2 *Let $t > 0$ and*

$$S_t(\lambda, A, \bar{f}) := \lambda \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{(n-1)!} e^{n\lambda t} R(n\lambda, A) \bar{f}.$$

Then for some $w \geq 0$ and $D \geq 1$, we have for all λ big enough and all $\bar{f} \in \mathcal{D}(A)$ that

$$\|\bar{S}_t \bar{f} - S_t(\lambda, A, \bar{f})\| \leq D \|A \bar{f}\| e^{wt} \lambda^{-1/2} + \exp(-e^{\lambda t}) \|\bar{f}\|.$$

Moreover, if the S_t form the strongly continuous transition semigroup of a Feller process, then

$$\|\bar{S}_t \bar{f} - S_t(\lambda, A, \bar{f})\| \leq D \|A \bar{f}\| \lambda^{-1/2} + \exp(-e^{\lambda t}) \|\bar{f}\|.$$

In the next result, we specialise to the situation where the semigroup acts on elements of a Hilbert space. So we assume that (S_t) is a Feller semigroup defined on a Hilbert space $\mathcal{H}$ with generator A. Let P_k be projections of $\mathcal{H}$ onto $\mathcal{H}_k$ with norm $\|P_k\| = 1$, typically orthogonal projections. Let A_k be as in (2.4), and let $R(\lambda, A)$ and $R(\lambda, A_k)$ be the corresponding resolvents. For $f^k \in \mathcal{D}(A_k)$, we consider $R(\lambda, A_k) \bar{f}^k \in \mathcal{H}_k$ as an element of $\mathcal{H}$.

Lemma A.3 *In the setting just described, assume that the operators $A_k : \mathcal{H}_k \to \mathcal{H}_k$ are such that with $\bar{f}^k = P_k \bar{f}$, $\bar{f} \in \mathcal{D}(A)$ and $\bar{f}^k \in \mathcal{D}(A_k)$, we have*

$$\lim_{k \to \infty} R(n\lambda, A_k) \bar{f}^k = R(n\lambda, A) \bar{f}$$

for all $n \geq 1$ and where the limit is taken in $\mathcal{H}$. Assume, too, that the norms of the semigroups (S_t^k) generated by the A_k as well as the norms of the (S_t) satisfy the bound in (3.2). Then $S_t(\lambda, A_k, \bar{f}^k) \to S_t(\lambda, A, \bar{f})$ for $k \to \infty$.

Proof From (3.2), it follows that $\|R(\lambda, A)\| \leq \frac{C}{\lambda - w}$, which is at most equal to $\frac{2C}{\lambda}$ for all $\lambda \geq 2w$. It follows that then $\|R(n\lambda, A)\bar{f}\| \leq \frac{2C}{n\lambda}\|\bar{f}\|$ whenever $n\lambda \geq 2w$. But by the same token, we also have that $\|R(n\lambda, A_k)\bar{f}^k\| \leq \frac{2C}{n\lambda}\|\bar{f}^k\| \leq \frac{2C}{n\lambda}\|\bar{f}\|$ since $\|\bar{f}^k\| \leq \|\bar{f}\|$.

Consider the norm of the summands in $S_t(\lambda, A_k, \bar{f}_k)$. For each n, this norm is at most

$$\lambda \frac{e^{n\lambda t}}{(n-1)!}\|R(n\lambda, A_k)\bar{f}^k\| \leq \lambda \frac{e^{n\lambda t}}{(n-1)!}\frac{2C}{n\lambda}\|\bar{f}\| = \frac{e^{n\lambda t}}{n!}\frac{2C}{\lambda}\|\bar{f}\|,$$

which has a finite sum over $n \geq 1$. Hence, considering the infinite sum

$$S_t(\lambda, A_k, \bar{f}_k) = \lambda \sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{(n-1)!}e^{n\lambda t}R(n\lambda, A_k)\bar{f}^k \tag{A.3}$$

as a Bochner integral, we can apply dominated convergence for Bochner integrals (see Hytönen et al. [17, Proposition 1.2.5]) to (A.3) to arrive at the convergence

$$\lambda \sum_{n=1}^{\infty}(-1)^{n-1}\frac{e^{n\lambda t}}{(n-1)!}R(n\lambda, A_k)\bar{f}^k \longrightarrow \lambda \sum_{n=1}^{\infty}(-1)^{n-1}\frac{e^{n\lambda t}}{(n-1)!}R(n\lambda, A)\bar{f},$$

which was our aim. □

Appendix B: Results on Hermite polynomials

This section proves the last equality in Lemma 3.10. The other results are standard.

Proof We compute the Fourier coefficient of $e^x \mathrm{He}_m^{\alpha,\beta}(x)$ and develop its orthogonal expansion with the Hermite polynomials as

$$e^x \mathrm{He}_m^{\alpha,\beta}(x) = \sum_{n=0}^{\infty} \frac{\int_{-\infty}^{\infty} e^x \mathrm{He}_n^{\alpha,\beta}(x)\mathrm{He}_m^{\alpha,\beta}(x)e^{-\frac{(x-\beta)^2}{2\alpha}}\mathrm{d}x}{\int_{-\infty}^{\infty}\mathrm{He}_n^{\alpha,\beta}(x)^2 e^{-\frac{(x-\beta)^2}{2\alpha}}\mathrm{d}x}\mathrm{He}_n^{\alpha,\beta}(x).$$

We start by simplifying the numerator. The product of two Hermite polynomials can be written as the sum of Hermite polynomials via

$$\mathrm{He}_n(x)\mathrm{He}_m(x) = \sum_{k=0}^{m\wedge n} k!\binom{m}{k}\binom{n}{k}\mathrm{He}_{m+n-2k}(x).$$

Furthermore, by applying the Hermite transform to e^{tx}, we have for any $k, t > 0$ that

$$\int_{-\infty}^{\infty} e^{tx}\mathrm{He}_k(x)e^{-\frac{x^2}{2}}\,\mathrm{d}x = t^k e^{\frac{t^2}{2}}\sqrt{2\pi}.$$

By substitution, we can calculate

$$\begin{aligned}\int_{-\infty}^{\infty} e^{tx}\mathrm{He}_n(x)\mathrm{He}_m(x)e^{-\frac{x^2}{2}}\,\mathrm{d}x &= \sum_{k=0}^{m\wedge n} k!\binom{m}{k}\binom{n}{k}\int_{-\infty}^{\infty} e^{tx}\mathrm{He}_{m+n-2k}(x)e^{-\frac{x^2}{2}}\,\mathrm{d}x\\ &= \sum_{k=0}^{m\wedge n} k!\binom{m}{k}\binom{n}{k}t^{m+n-2k}e^{\frac{t^2}{2}}\sqrt{2\pi}.\end{aligned}$$

Using integration by substitution, we get

$$\begin{aligned}&\int_{-\infty}^{\infty} e^{x}\mathrm{He}_n^{\alpha,\beta}(x)\mathrm{He}_m^{\alpha,\beta}(x)e^{-\frac{(x-\beta)^2}{2\alpha}}\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} e^{x}\alpha^{\frac{n}{2}}\mathrm{He}_n\left(\frac{x-\beta}{\sqrt{\alpha}}\right)\alpha^{\frac{m}{2}}\mathrm{He}_m\left(\frac{x-\beta}{\sqrt{\alpha}}\right)e^{-\frac{(x-\beta)^2}{2\alpha}}\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} e^{\sqrt{\alpha}y+\beta}\alpha^{\frac{n+m+1}{2}}\mathrm{He}_n(y)\mathrm{He}_m(y)e^{-\frac{y^2}{2}}\,\mathrm{d}y\\ &= e^{\beta+\frac{\alpha}{2}}\sqrt{2\pi\alpha}\sum_{k=0}^{m\wedge n} k!\binom{m}{k}\binom{n}{k}\alpha^{m+n-k}.\end{aligned}$$

In simplifying the denominator, we start from

$$\int_{-\infty}^{\infty}\mathrm{He}_n(x)^2e^{-\frac{x^2}{2}}\,\mathrm{d}x = \sqrt{2\pi}n!.$$

By the same substitution argument, we get

$$\int_{-\infty}^{\infty}\mathrm{He}_n^{\alpha,\beta}(x)^2e^{-\frac{(x-\beta)^2}{2\alpha}}\,\mathrm{d}x = \alpha^n\sqrt{2\alpha\pi}n!.$$

Finally, substituting the solved integrals in the numerator and denominator into the original equation gives

$$\begin{aligned}e^{x}\mathrm{He}_m^{\alpha,\beta}(x) &= \sum_{n=0}^{\infty}\frac{e^{\beta+\frac{\alpha}{2}}\sqrt{2\pi\alpha}\sum_{k=0}^{m\wedge n}k!\binom{m}{k}\binom{n}{k}\alpha^{m+n-k}}{\alpha^n\sqrt{2\alpha\pi}n!}\mathrm{He}_n^{\alpha,\beta}(x)\\ &= \sum_{n=0}^{\infty}e^{\beta+\frac{\alpha}{2}}\sum_{k=0}^{m\wedge n}\binom{m}{k}\frac{\alpha^{m-k}}{(n-k)!}\mathrm{He}_n^{\alpha,\beta}(x).\end{aligned}$$

□

Acknowledgements We express our gratitude to Richard Martin, Andrew Ang, Fabio Mercurio and Yury Krongauz for valuable input on our paper, including the suggested application to Gram–Charlier series. We also thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Aït-Sahalia, Y.: Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica **70**, 223–262 (2002)
2. Al-Mohy, A.H., Higham, N.J.: Computing the action of the matrix exponential, with an application to exponential integrators. SIAM J. Sci. Comput. **33**, 488–511 (2011)
3. Antonov, A., Spector, M.: General short-rate analytics. Risk **24**(5), 66–71 (2011)
4. Arendt, W., Batty, C.J.K., Hieber, M., Neubrander, F.: Vector-Valued Laplace Transforms and Cauchy Problems, 2nd edn. Springer, Basel (2011)
5. Arvanitis, A., Gregory, J., Laurent, J.-P.: Building models for credit spreads. J. Deriv. **6**(3), 27–43 (1999)
6. Ba, M., van Beek, M., Zhao, C.: An integrated credit model. Working paper, BlackRock, New York (2020)
7. Black, F., Karasinski, P.: Bond and option pricing when short rates are lognormal. Financ. Anal. J. **47**(4), 52–59 (1991)
8. Bobrowski, A.: Functional Analysis for Probability and Stochastic Processes: An Introduction. Cambridge University Press, Cambridge (2005)
9. Böttcher, A., Wenzel, D.: How big can the commutator of two matrices be and how big is it typically? Linear Algebra Appl. **403**, 216–228 (2005)
10. Capriotti, L.: The exponent expansion: an effective approximation of transition probabilities of diffusion processes and pricing kernels of financial derivatives. Int. J. Theor. Appl. Finance **9**, 1179–1199 (2006)
11. Chateau, J.-P., Dufresne, D.: Gram-Charlier processes and applications to option pricing. J. Probab. Stat. **2017**, Article ID 8690491-1–19 (2017)
12. Cox, J.C., Ingersoll, J.E. Jr., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**, 385–407 (1985)
13. Cuchiero, C., Keller-Ressel, M., Teichmann, J.: Polynomial processes and their applications to mathematical finance. Finance Stoch. **16**, 711–740 (2012)
14. Duffie, D., Filipović, D., Schachermayer, W.: Affine processes and applications in finance. Ann. Appl. Probab. **13**, 984–1053 (2003)
15. Hagen, R., Roch, S., Silbermann, B.: C^*-Algebras and Numerical Analysis. CRC Press, Boca Raton (2000)
16. Hurd, T., Kuznetsov, A.: Affine Markov chain models of multifirm credit migration. J. Credit Risk **3**(1), 3–29 (2007)
17. Hytönen, T., Van Neerven, J., Veraar, M., Weis, L.: Analysis in Banach Spaces Volume I: Martingales and Littlewood–Paley Theory. Springer, Berlin (2016)
18. Jarrow, R.A., Lando, D., Turnbull, S.M.: A Markov model for the term structure of credit risk spreads. Rev. Financ. Stud. **10**, 481–523 (1997)
19. Lando, D.: On Cox processes and credit risky securities. Rev. Deriv. Res. **2**, 99–120 (1998)
20. Li, M., Mercurio, F., Resnick, S.: The GARCH linear SDE: explicit formulas and the pricing of a quanto CDS. Preprint (2018). https://ssrn.com/abstract=3188272

21. Lindner, M.: Fredholm theory and stable approximation of band operators and their generalisations. PhD Thesis, Technische Universität Chemnitz (2009). https://nbn-resolving.org/urn:nbn:de:bsz:ch1-200901182
22. Martin, R.J.: Credit migration: Generating generators. Preprint (2020). https://arxiv.org/abs/2006.11146
23. Neubrander, F.: On the relation between the semigroup and its infinitesimal generator. Proc. Am. Math. Soc. **100**, 104–108 (1987)
24. Popovic, R., Goldsman, D.: Easy Gram–Charlier valuations of options. J. Deriv. **20**(2), 79–97 (2012)
25. Rabinovich, V.S., Roch, S., Silbermann, B.: On finite sections of band-dominated operators. In: Bastos, M.A., et al. (eds.) Operator Algebras, Operator Theory and Applications, pp. 385–391. Birkhäuser, Basel (2008)
26. Robert, L., Santiago, L.: Finite sections method for Hessenberg matrices. J. Approx. Theory **123**, 68–88 (2003)
27. Stehlíková, B., Capriotti, L.: An effective approximation for zero-coupon bonds and Arrow–Debreu prices in the Black–Karasinski model. Int. J. Theor. Appl. Finance **17**, 1450037 (2014)
28. Tanaka, K., Yamada, T., Watanabe, T.: Applications of Gram–Charlier expansion and bond moments for pricing of interest rates and credit risk. Quant. Finance **10**, 645–662 (2010)
29. Tourrucôo, F., Hagan, P.S., Schleiniger, G.F.: Approximate formulas for zero-coupon bonds. Appl. Math. Finance **14**, 207–226 (2007)
30. Zhou, H.: Itô conditional moment generator and the estimation of short-rate processes. J. Financ. Econom. **1**, 250–271 (2003)