

- **Calibration of local correlation models to basket smiles**
Julien Guyon
- **Error analysis in Fourier methods for option pricing**
Fabián Crocce, Juho Häppälä, Jonas Kiesling and Raúl Tempone
- **Efficient estimation of sensitivities for counterparty credit risk with the finite difference Monte Carlo method**
Cornelis S. L. de Graaf, Drona Kandhai and Peter M. A. Sloot
- **Smile with the Gaussian term structure model**
Abdelkoddousse Ahdida, Aurélien Alfonsi and Ernesto Palidda

The Journal of

Computational Finance

For all subscription queries, please call:

UK/Europe: +44 (0) 207 316 9300

USA: +1 646 736 1850 ROW: +852 3411 4828

Risk.net in numbers

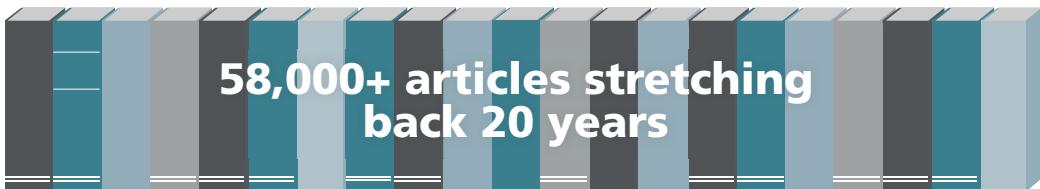


New articles &
technical papers
(each month)

140,000



Users
(each month)



6,500+
on Asset Management



6,900+
on Commodities



21,000+
on Derivatives



19,400+
on Regulation



19,600+
on Risk Management

See what you're missing

Visit the world's leading source of exclusive in-depth news & analysis on risk management, derivatives and complex finance now.

The Journal of Computational Finance

EDITORIAL BOARD

Editor-in-Chief

CORNELIS W. OOSTERLEE CWI – Dutch Center for Mathematics and Computer Science, Amsterdam

Editor Emeritus

MARK BROADIE Columbia University

Associate Editors

LEIF ANDERSEN Bank of America

Merrill Lynch

PETER CARR Morgan Stanley

DUY-MINH DANG University of Queensland

M. A. H. DEMPSTER University of Cambridge

PETER FORSYTH University of Waterloo

MIKE GILES Oxford University

JONATHAN GOODMAN New York University

LUIS ORTIZ GRACIA The University of

Barcelona

LECH A. GRZELAK Rabobank International

DESMOND J. HIGHAM University of Strathclyde

KAREL IN 'T HOUT University of Antwerp

YUYING LI University of Waterloo

ANDREW LO Massachusetts Institute of
Technology

MIKE LUDKOVSKI University of California
Santa Barbara

ANDREA PASCUCCI University of Bologna

VLADIMIR V. PITERBARG Rokos

CHRISTOPH REISINGER Oxford University

CHRIS ROGERS University of Cambridge

JOHN SCHOENMAKERS Weierstrass Institute,
Berlin

ARTUR SEPP Julius Baer

KENNETH SINGLETON Stanford University

REHA TUTUNCU Goldman Sachs, New York

CARLOS VÁZQUEZ CENDÓN University of
La Coruña

KENNETH VETZAL University of Waterloo

NANCY WALLACE University of California,
Berkeley

SUBSCRIPTIONS

The Journal of Computational Finance (Print ISSN 1460-1559 | Online ISSN 1755-2850) is published quarterly by Infopro Digital, Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK. Subscriptions are available on an annual basis, and the rates are set out in the table below.

	UK	Europe	US
Risk.net Journals	£2045	€2895	\$3225
Individual Journal (print)	£775	€1075	\$1265
Risk.net Premium	£2895	€4145	\$4545

Academic discounts are available. Please enquire by using one of the contact methods below.

All prices include postage. All subscription orders, single/back issues orders, and changes of address should be sent to:

UK & Europe Office: Infopro Digital, Haymarket House, 28-29 Haymarket, London, SW1Y 4RX, UK. Tel: +44 207 316 9300

US & Canada Office: Infopro Digital, 55 Broad Street, Floor 22, New York, NY, 10005.
Tel: +1 646 736 1850; Fax: +1 646-390-6612

Asia & Pacific Office: Infopro Digital, Unit 1704-05, Berkshire House, Taikoo Place, 25 Westlands Road, Quarry Bay, Hong Kong. Tel: +852 3411 4828 Fax: +852 3411 4811 Website:
www.risk.net/journal E-mail: info@risk.net

The Journal of Computational Finance

GENERAL SUBMISSION GUIDELINES

Manuscripts and research papers submitted for consideration must be original work that is not simultaneously under review for publication in another journal or other publication outlets. All articles submitted for consideration should follow strict academic standards in both theoretical content and empirical results. Articles should be of interest to a broad audience of sophisticated practitioners and academics.

Submitted papers should follow *Webster's New Collegiate Dictionary* for spelling, and *The Chicago Manual of Style* for punctuation and other points of style, apart from a few minor exceptions that can be found at www.risk.net/journal. Papers should be submitted electronically via email to: journals@incisivemedia.com. Please clearly indicate which journal you are submitting to.

You must submit two versions of your paper; a single L^AT_EX version and a PDF file. L^AT_EX files need to have an explicitly coded bibliography included. All files must be clearly named and saved by author name and date of submission. All figures and tables must be included in the main PDF document and also submitted as separate editable files and be clearly numbered.

All papers should include a title page as a separate document, and the full names, affiliations and email addresses of all authors should be included. A concise and factual abstract of between 150 and 200 words is required and it should be included in the main document. Five or six keywords should be included after the abstract. Submitted papers must also include an Acknowledgements section and a Declaration of Interest section. Authors should declare any funding for the article or conflicts of interest. Citations in the text must be written as (John 1999; Paul 2003; Peter and Paul 2000) or (John *et al* 1993; Peter 2000).

The number of figures and tables included in a paper should be kept to a minimum. Figures and tables must be included in the main PDF document and also submitted as separate individual editable files. Figures will appear in color online, but will be printed in black and white. Footnotes should be used sparingly. If footnotes are required then these should be included at the end of the page and should be no more than two sentences. Appendixes will be published online as supplementary material.

Before submitting a paper, authors should consult the full author guidelines at:

<http://www.risk.net/static/risk-journals-submission-guidelines>

Queries may also be sent to:

The Journal of Computational Finance, Infopro Digital,
Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK
Tel: +44 (0)20 7004 7531; Fax: +44 (0)20 7484 9758
E-mail: journals@incisivemedia.com

To subscribe to a Risk Journals visit subscriptions.risk.net/journals or email info@risk.net

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net

The Journal of Computational Finance

GENERAL SUBMISSION GUIDELINES

Manuscripts and research papers submitted for consideration must be original work that is not simultaneously under review for publication in another journal or other publication outlets. All articles submitted for consideration should follow strict academic standards in both theoretical content and empirical results. Articles should be of interest to a broad audience of sophisticated practitioners and academics.

Submitted papers should follow *Webster's New Collegiate Dictionary* for spelling, and *The Chicago Manual of Style* for punctuation and other points of style, apart from a few minor exceptions that can be found at www.risk.net/journal. Papers should be submitted electronically via email to: journals@incisivemedia.com. Please clearly indicate which journal you are submitting to.

You must submit two versions of your paper; a single L^AT_EX version and a PDF file. L^AT_EX files need to have an explicitly coded bibliography included. All files must be clearly named and saved by author name and date of submission. All figures and tables must be included in the main PDF document and also submitted as separate editable files and be clearly numbered.

All papers should include a title page as a separate document, and the full names, affiliations and email addresses of all authors should be included. A concise and factual abstract of between 150 and 200 words is required and it should be included in the main document. Five or six keywords should be included after the abstract. Submitted papers must also include an Acknowledgements section and a Declaration of Interest section. Authors should declare any funding for the article or conflicts of interest. Citations in the text must be written as (John 1999; Paul 2003; Peter and Paul 2000) or (John *et al* 1993; Peter 2000).

The number of figures and tables included in a paper should be kept to a minimum. Figures and tables must be included in the main PDF document and also submitted as separate individual editable files. Figures will appear in color online, but will be printed in black and white. Footnotes should be used sparingly. If footnotes are required then these should be included at the end of the page and should be no more than two sentences. Appendixes will be published online as supplementary material.

Before submitting a paper, authors should consult the full author guidelines at:

<http://www.risk.net/static/risk-journals-submission-guidelines>

Queries may also be sent to:

The Journal of Computational Finance, Infopro Digital,
Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK
Tel: +44 (0)20 7004 7531; Fax: +44 (0)20 7484 9758
E-mail: journals@incisivemedia.com

The Journal of

Computational Finance

The journal

The Journal of Computational Finance welcomes papers dealing with innovative computational techniques in the following areas.

- Numerical solutions of pricing equations: finite differences, finite elements, and spectral techniques in one and multiple dimensions.
 - Simulation approaches in pricing and risk management: advances in Monte Carlo and quasi-Monte Carlo methodologies; new strategies for market factors simulation.
 - Optimization techniques in hedging and risk management.
 - Fundamental numerical analysis relevant to finance: effect of boundary treatments on accuracy; new discretization of time-series analysis.
 - Developments in free-boundary problems in finance: alternative ways and numerical implications in American option pricing.
-

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net

CONTENTS

Letter from the Editor-in-Chief	vii
RESEARCH PAPERS	
<i>Calibration of local correlation models to basket smiles</i>	1
Julien Guyon	
<i>Error analysis in Fourier methods for option pricing</i>	53
Fabián Crocce, Juho Häppölä, Jonas Kiessling and Raúl Tempone	
<i>Efficient estimation of sensitivities for counterparty credit risk with the finite difference Monte Carlo method</i>	83
Cornelis S. L. de Graaf, Drona Kandhai and Peter M. A. Sloot	
<i>Smile with the Gaussian term structure model</i>	115
Abdelkoddousse Ahdida, Aurélien Alfonsi and Ernesto Palidda	

Editor-in-Chief: Cornelis W. Oosterlee
Publisher: Nick Carver
Journals Manager: Dawn Hunter
Editorial Assistant: Carolyn Moclair

Subscription Sales Manager: Aaraa Javed
Global Key Account Sales Director: Michelle Godwin
Composition and copyediting: T&T Productions Ltd
Printed in UK by Printondemand-Worldwide

©Copyright Infopro Digital Limited, 2017. All rights reserved. No parts of this publication may be reproduced, stored in or introduced into any retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the copyright owners.



To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net

LETTER FROM THE EDITOR-IN-CHIEF

Cornelis W. Oosterlee

CWI – Dutch Center for Mathematics and Computer Science, Amsterdam

The June 2017 *Journal of Computational Finance* is, again, an interesting issue in terms of the numerical techniques its authors employ. However, before we launch into an overview of the featured papers, I would like to welcome Professor Andrea Pascucci, from the University of Bologna, Italy, to our editorial board. Professor Pascucci is an expert on, among other topics, the derivation of asymptotic formulas and corresponding efficient valuation techniques when dealing with involved processes in computational finance. It is great to have someone with his level of expertise on board.

The current issue is, as usual, diverse in the numerical techniques it presents, and modern in terms of the financial applications it considers. We encounter a Monte Carlo type particle method in the context of a local correlation model; error analysis for Fourier type option pricing methods; a finite difference partial differential equation (PDE) method in the context of credit valuation adjustments (CVAs); and a comparison of different numerical methods in the context of a Gaussian term structure model.

We begin with “Calibration of local correlation models to basket smiles”. Here, Julien Guyon from Bloomberg L.P. introduces a new family of local correlation models in order to accurately calibrate to the implied volatility smile of basket options, which may arise in a variety of asset classes. State-dependent correlations allow for better hedging. By incorporating this model into the framework of the Monte Carlo particle method, calibrated local correlation models can be built. The technique is generalized to models based on stochastic interest rates, stochastic dividend yields, local stochastic volatility and local correlation.

The second contribution to this issue is “Error analysis in Fourier methods for option pricing” by Fabián Crocce, Juho Häppölä, Jonas Kiesling and Raúl Tempone. In this paper the errors made when using a Fourier method to price European options under exponential Lévy processes are analyzed and bounded. The bound is analyzed in detail, and it is also used to determine different method parameters for the numerical technique to converge in a robust and efficient way.

Our third paper is “Efficient estimation of sensitivities for counterparty credit risk with the finite difference Monte Carlo method” by Cornelis S. L. de Graaf, Drona Kandhai and Peter M. A. Sloot. In this paper Monte Carlo forward asset path generation is combined with the numerical solution of a PDE to compute the exposure and potential future exposure related to the option contract by an iteration backward

in time. The asset model considered is based on stochastic volatility and a stochastic interest rate, which results in a three-dimensional PDE (plus time as the fourth problem dimension). The proposed finite difference Monte Carlo method is particularly useful for the efficient computation of CVA sensitivities.

In “Smile with the Gaussian term structure model”, Abdelkoddousse Ahdida, Aurélien Alfonsi and Ernesto Palidda introduce an affine extension of the linear Gaussian term structure model that gives rise to an implied volatility smile. A Wishart-driven model for interest rates is proposed, in which the parameters and state variables may be clearly interpreted with regard to the corresponding yield curve dynamics. Different numerical models have been compared with this model; these range from Fourier and Laplace transform based methods via perturbation theory based efficient pricing methods to a second-order stochastic differential equation discretization scheme, which forms the basis of a Monte Carlo method.

I wish you very enjoyable reading of this issue of *The Journal of Computational Finance*.



Research Paper

Calibration of local correlation models to basket smiles

Julien Guyon

¹Quantitative Research, Bloomberg L.P., 731 Lexington Avenue,
New York, NY 10022, USA; email: jguyon2@bloomberg.net

²Department of Mathematics, Columbia University, Room 509, MC 4406,
2990 Broadway, New York, NY 10027, USA; email: jg3601@columbia.edu

³Courant Institute of Mathematical Sciences, NYU, 251 Mercer Street,
New York, NY 10012, USA; email: julien.guyon@nyu.edu

(Received November 30, 2013; revised July 23, 2015; accepted September 15, 2015)

ABSTRACT

Allowing correlation to be local, ie, state-dependent, in multi-asset models allows better hedging by incorporating correlation moves in the Delta. When options on a basket – be it a stock index, a cross-foreign exchange rate or an interest rate spread – are liquidly traded, one may want to calibrate a local correlation to these option prices. Only two particular solutions have been suggested so far in the literature. Both impose a particular dependency of the correlation matrix on the asset values that one has no reason to undergo. They may also fail to be admissible, ie, positive semi-definite. We explain how, by combining the particle method presented in “The smile calibration problem solved” by Guyon and Henry-Labordère (2011) with a simple affine transform, we can build all the calibrated local correlation models. The two existing models appear as special cases (if admissible). For the first time, one can now choose a calibrated local correlation in order to fit a view on the correlation skew, or reproduce historical correlation, or match some exotic option prices, thus improving the pricing, hedging and risk-management of multi-asset derivatives. This technique is generalized to calibrate models that combine stochastic interest rates,

stochastic dividend yield, local stochastic volatility and local correlation. Numerical results show the wide variety of calibrated local correlations and give insight into a difficult (still unsolved) problem: finding lower bounds/upper bounds on general multi-asset option prices given the whole surfaces of implied volatilities of a basket and its constituents.

Keywords: local correlation; calibration; particle method; McKean nonlinear stochastic differential equation (SDE); path-dependent correlation.

1 INTRODUCTION

Many practitioners use a multi-asset version of the local volatility model (Dupire 1994) to price multi-asset derivatives. Most of the time, the correlation matrix is assumed to be constant, eg, some constant historical correlation ρ^{hist} . In the equity market, since banks usually “sell correlation”, ie, sell products that have a positive sensitivity to correlation, they tend to overprice correlation. They often use a convex combination of ρ^{hist} and the matrix $\mathbf{1}$, representing full correlation of the assets, whose entries are all equal to 1:

$$\rho = (1 - \lambda)\rho^{\text{hist}} + \lambda\mathbf{1}, \quad \lambda \in [0, 1].$$

However, such a model is not able to reproduce the market smile of implied volatilities of stock index options. Typically, when a constant correlation is picked to match the price of the at-the-money implied volatility of the index, it generates a skew that is much smaller (roughly two times smaller) than the market skew (see Bakshi *et al* 2003; Bollen and Whaley 2004; Branger and Schlag 2004). In other words, assuming single-asset local volatilities for the stocks, which we will do throughout this paper, the smile of index options contains information on how its constituents are more correlated in a bearish market and less correlated in a bullish one.

Local correlation models, in which the correlation matrix is allowed to be state-dependent,

$$\rho(t, S_t^1, \dots, S_t^N),$$

are able to capture this information. They are of high practical importance, not only because they include correlation variability in option prices, but mainly because they allow better hedging by incorporating correlation moves in the Delta. This is crucial for short cross-Gamma positions, where an underestimation of correlation in periods of crisis yields a daily profit-and-loss (P&L) bleeding that can only be stopped by incurring a large remarking-to-market loss. Many investment banks were affected by this in 2008 after the bankruptcy of Lehman Brothers. Like the local volatility model, local correlation models do not aim at describing the real-world dynamics of

the assets; rather, they aim at helping traders risk-manage their correlation positions, especially during crises. Local correlation models are also very useful in the context of foreign exchange (FX) options. They allow us to build models that are consistent with the market smiles of two FX rates as well as the market smile of the cross-rate. They are also used in interest rates to calibrate to spread option prices.

Let us say that a local correlation model is admissible if it calibrates to the market smile of the index. To the best of our knowledge, only two admissible models have been suggested so far in the literature. In both models, the correlation matrix $\rho = (1 - \lambda)\rho^0 + \lambda\rho^1$ is assumed to lie on the line defined by two fixed correlation matrixes ρ^0 and ρ^1 . The first model, proposed by Langnau (2010), assumes that the instantaneous variance of a stock index in the multi-asset local volatility–local correlation model is local in index, ie, depends on the stocks only through the index value. The second model, presented in Reghai (2010) and Guyon and Henry-Labordère (2011b), assumes that the instantaneous correlation itself (or equivalently λ) is local in index. It may seem enough to have those two models at hand. However, they both have drawbacks. First, both models may actually fail to be admissible. In Langnau (2010) and Guyon and Henry-Labordère (2011b), a unique correlation candidate is explicitly built and may fail to be positive semi-definite (PSD). Then, one projects the candidate onto the set of correlation matrixes, and the resulting model does not perfectly calibrate. The correlation candidate has a greater chance of being PSD in the second model (see (6.1)). In Reghai (2010), the correlation matrix is built by solving a fixed point problem that may have no solution. (More precisely, it has no solution when the correlation candidate exhibited in Guyon and Henry-Labordère (2011b) is not PSD.) Second, even if both models are admissible, there is no reason why one would undergo either correlation structure. For instance, the resulting correlation may have a weird skew (dependence on the asset values), or its skew may be far from that which is historically observed, or it may generate prices of exotic options that are far from market quotes.

In this paper, we build a whole family of local correlation models by combining the particle method with a new, simple idea. This family is parameterized by two functions a and b that depend on time and on the values of all the underlying assets. Instead of assuming that the basket variance or the correlation (or equivalently λ) is local in index, we assume this of $a + b\lambda$. This new, simple idea may seem ad hoc at first, but it actually generates all the calibrated models such that $\rho \in (\rho^0, \rho^1)$. In particular, by allowing ρ^0 and ρ^1 to be state-dependent and varying them, we can span all the admissible local correlations. The two existing models are just two particular elements in this family: they correspond to two particular choices of (a, b) . We easily handle path-dependent correlation by also allowing a and b to depend on any set of path-dependent variables. Table 1 compares the two existing models with the new family of models.

TABLE 1 Summary of models and methods for calibrating to basket smile.

	Langnau (2010)	Reghai (2010)	Guyon and Henry-Labordère (2011b)	New model
Function of basket value	Basket variance	Correlation, λ	Correlation, λ	$a + b\lambda$
Function of all underlying assets	Correlation, λ	Basket variance	Basket variance	a, b , basket variance and correlation
Possibly function of any path-dependent variable	—	—	—	a, b , basket variance and correlation
Calibration method	Closed form	Fixed point	Particle method	Particle method
Correlation candidate built explicitly	Yes	No	Yes, time step by time step	Yes, time step by time step
Avoids computing implied volatilities	Yes	No	Yes	Yes
Number of degrees of freedom	0	0	0	An infinity: all functions a and b

The correlation matrix $\rho = (1 - \lambda)\rho^0 + \lambda\rho^1$ lies on the line defined by two fixed correlation matrixes ρ^0 and ρ^1 .

Using the new family, we can now design our favorite local correlation model in order to satisfy desirable properties \mathcal{P} , such as matching a view on a correlation skew, reproducing some features of historical correlation, calibrating to other option prices, etc, on top of reproducing the market smile of the basket, be it a stock index, a cross-FX rate, an interest rate spread, etc. Not all models in the family are admissible: for a given (a, b) , the particle method generates an explicit local correlation candidate. Admissible models correspond to those pairs (a, b) for which the candidate is PSD at all times and for all asset values. Otherwise, we may project the candidate onto the set of correlation matrixes and carry on using the particle method. In those cases, the resulting model is not perfectly admissible, but it may be accurate enough for trading purposes: for instance, when the correlation candidate fails to be PSD only for unlikely asset values (see examples in Section 9).

A summarized version of this paper was published in *Risk* (Guyon 2014a). Relaxing the assumption on volatilities is another way to generate steep index skews. This is explained in Guyon (2016), which introduces “cross-dependent volatility models”, ie, multi-asset models in which the volatility of each asset is a function of not only its current level but also those of the other assets. Another way of jointly calibrating to the smile of a basket and to the smiles of its components is described in Jourdain and Sbai (2012). Here, the authors build an incomplete stochastic volatility–stochastic correlation model by following a top-down approach, in which the level of a stock index induces some feedback on the dynamics of its constituents. Attempts to approximately calibrate to a triangle of FX market smiles in a symmetric way include De Col *et al* (2013), in which a multi-Heston model with constant correlations is used. Reghai (2010) considers the pricing of worst-of options in a model, where the local correlation depends on the stocks only through the worst performance of the basket constituents, and suggests a historical calibration procedure. Delanoe (2013) addresses the question of calibrating such a model to option prices and discusses stochastic volatility extensions of local correlation models. In the context of constant correlation, Avellaneda *et al* (2002) were the first to give the formula for the equivalent local volatility of a basket of stocks (see (8.2)) and estimate it using short-term asymptotics at order zero, namely via Varadhan’s formula and the method of steepest descent. The expansion at order one, as well as an extension to local in index correlation models, is proved in Henry-Labordère (2009). Durrelman and El Karoui (2008) price options written on a domestic asset based on implied volatilities of options on the same asset expressed in a foreign currency and the exchange rate. Given a local correlation, they derive explicit formulas to compute the at-the-money implied volatility, skew, convexity and term structure for short maturities. In Cont and Deguest (2010), the authors use a random mixture of reference models to build a multi-asset model consistent with a set of observed single- and multi-asset derivative prices. Austing (2011) provides an analytic formula for a joint probability density such that all three market smiles

in an FX triangle are repriced. A few stochastic correlation models have also been suggested and analyzed in the literature, including those by Gourieroux and Sufana (2003), Da Fonseca *et al* (2008) and Ahdida and Alfonsi (2013).

This paper is structured as follows. In Section 3, we introduce our new family of local correlation models in the simple context of the FX triangle smile calibration problem, which is briefly recalled in Section 2. Then, in Section 4, we show how the family is built step by step, from inception to maturity, using the particle method. We highlight some key examples in Section 5. Important links between the various admissible local correlations are investigated in Section 6. Section 7 deals with the impact of correlation on the prices of multi-asset options, with a reminder about implied correlation *à la* Dupire (1998) and a new formula *à la* Gatheral (2006). Section 8 shows how we extend our method to the N -dimensional stock index smile calibration problem. In Section 9, our numerical examples, both in the FX and the equity contexts, show the wide variety of admissible correlations and give insight into lower bounds and upper bounds on prices of multi-asset options when the smile of a basket and the smiles of its constituents are given. In Section 10, we generalize our results to models that combine stochastic interest rates, stochastic dividend yield, local stochastic volatility and local correlation. Finally, we conclude in Section 11. The proofs are gathered in the online appendix.

2 THE FOREIGN EXCHANGE TRIANGLE SMILE CALIBRATION PROBLEM

Let us introduce our new family of local correlation models in the simple context of the FX triangle smile calibration problem. Section 8 deals with the general N -dimensional basket case. Let S^1, S^2 be two FX rates, and let $S^{12} = S^1/S^2$ be the cross-rate. One can think of $S^1 = \text{EUR/USD}$, $S^2 = \text{GBP/USD}$ and $S^{12} = \text{EUR/GBP}$. Assume we know from the market the surfaces of implied volatility for S^1 , S^2 and S^{12} until some maturity T , and that those surfaces are jointly arbitrage-free. They correspond to three local volatility surfaces that we denote by $\sigma_1(t, S^1)$, $\sigma_2(t, S^2)$ and $\sigma_{12}(t, S^{12})$. Assume the following model \mathcal{M}_ρ for the dynamics of S^1 and S^2 :

$$\begin{aligned} dS_t^1 &= (r_t^d - r_t^1)S_t^1 dt + \sigma_1(t, S_t^1)S_t^1 dW_t^1, \\ dS_t^2 &= (r_t^d - r_t^2)S_t^2 dt + \sigma_2(t, S_t^2)S_t^2 dW_t^2, \\ d\langle W^1, W^2 \rangle_t &= \rho(t, S_t^1, S_t^2) dt. \end{aligned} \tag{2.1}$$

All interest rates are deterministic, both S^1 and S^2 follow local volatility dynamics and the two driving processes W^1 and W^2 are Brownian motions under the risk-neutral measure \mathbb{Q} associated with the anchor (domestic) currency (USD in our example).

They have a local instantaneous correlation $\rho \in \mathcal{C}$, where \mathcal{C} denotes the set of functions $\rho: [0, T] \times \mathbb{R}_+^* \times \mathbb{R}_+^* \rightarrow [-1, 1]$.

Let $\mathbb{E}^{\mathbb{Q}^f}$ denote the expectation under the risk-neutral measure \mathbb{Q}^f associated with the foreign currency in S^2 (GBP in our example):

$$\frac{d\mathbb{Q}^f}{d\mathbb{Q}} = \frac{S_T^2}{S_0^2} \exp\left(\int_0^T (r_t^2 - r_t^d) dt\right). \quad (2.2)$$

Then we have the following (see proof in the online appendix).

PROPOSITION 2.1 *Model M_ρ is calibrated to the market smile of the cross-rate S^{12} if and only if, for all $t \in [0, T]$,*

$$\begin{aligned} \mathbb{E}_\rho^{\mathbb{Q}^f} \left[\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) - 2\rho(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2) \mid \frac{S_t^1}{S_t^2} \right] \\ = \sigma_{12}^2 \left(t, \frac{S_t^1}{S_t^2} \right). \end{aligned} \quad (2.3)$$

Equation (2.3) is equivalent to

$$\begin{aligned} \frac{\mathbb{E}_\rho^{\mathbb{Q}}[S_t^2(\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) - 2\rho(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)) \mid S_t^1/S_t^2]}{\mathbb{E}_\rho^{\mathbb{Q}}[S_t^2 \mid S_t^1/S_t^2]} \\ = \sigma_{12}^2 \left(t, \frac{S_t^1}{S_t^2} \right). \end{aligned} \quad (2.4)$$

Note that the left-hand side of (2.3) depends on the correlation in two ways: (i) explicitly through the random variable $\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$ and (ii) implicitly through the conditional distribution of (S_t^1, S_t^2) , given S_t^1/S_t^2 under \mathbb{Q}^f . To emphasize point (ii), we have written $\mathbb{E}_\rho^{\mathbb{Q}^f}$ instead of $\mathbb{E}^{\mathbb{Q}^f}$.

DEFINITION 2.2 Any $\rho \in \mathcal{C}$ satisfying (2.3) is called an admissible correlation.

In this paper, we are mainly interested in the following important practical question.

(Q) How do we build all the admissible correlations?

REMARK 2.3 We may have started with a general stochastic process (ρ_t) for the correlation, which possibly depends on extra sources of randomness. In this situation, the calibration condition (2.3) still holds with

$$\rho(t, S_t^1, S_t^2) \equiv \mathbb{E}_{\rho_t}^{\mathbb{Q}^f} [\rho_t \mid S_t^1, S_t^2] = \mathbb{E}_{\rho_t}^{\mathbb{Q}} [\rho_t \mid S_t^1, S_t^2].$$

This result is not completely trivial, as one needs to show that $\mathbb{E}_{\rho_t}^{\mathbb{Q}^f}$ can be replaced by $\mathbb{E}_\rho^{\mathbb{Q}^f}$. This follows from Gyöngy's theorem (see Section 7.2 for a simple derivation). As a consequence, if there exists an admissible correlation process ρ_t , then there exists an admissible local correlation $\rho(t, S^1, S^2)$. As far as calibration to the market smile of the cross-rate is concerned, assuming a local correlation $\rho(t, S^1, S^2)$ is not restrictive.

REMARK 2.4 A preliminary important and difficult question is this: how do we verify that the three surfaces of implied volatility for S^1 , S^2 and S^{12} are jointly arbitrage-free? How do we detect joint arbitrages? To the best of our knowledge, no explicit solution to this problem has been found so far. Partial results exist: Exercise 12.13.1 in Guyon and Henry-Labordère (2014) deals with the simpler case where only one maturity T is considered, while Exercise 12.13.2 deals with the case where finitely many maturities T are considered.

3 A NEW REPRESENTATION OF ADMISSIBLE CORRELATIONS

To answer question (Q), we now introduce a new representation of admissible correlations. The idea is to consider all the affine transforms $a(t, S^1, S^2) + b(t, S^1, S^2)\rho(t, S^1, S^2)$ of the local correlation that are functions of $(t, S^1/S^2)$ only, or, in other words, to write a local correlation function $\rho(t, S_1, S_2)$ as an affine transform $\alpha(t, S^1, S^2) + \beta(t, S^1, S^2)f(t, S^1/S^2)$ of a function of $(t, S^1/S^2)$ only.

DEFINITION 3.1 We say that a function is “local in X ” if it is a function of (t, X) only, say, $f(t, X)$. When $X = S^1/S^2$, we also say “local in cross”.

PROPOSITION 3.2 (Local in cross affine transform representation of admissible correlations) *Let $\rho \in \mathcal{C}$. It is an admissible correlation if and only if there exist two functions $a(t, S^1, S^2)$ and $b(t, S^1, S^2)$ such that b does not vanish and¹*

$$\begin{aligned} & \rho(t, S_t^1, S_t^2) \\ &= \frac{1}{b(t, S_t^1, S_t^2)} \\ & \times \left(\frac{\mathbb{E}_\rho^{\mathbb{Q}^f} [\sigma_1^2 + \sigma_2^2 + 2(a/b)\sigma_1\sigma_2 \mid S_t^1/S_t^2] - \sigma_{12}^2(t, S_t^1/S_t^2)}{2\mathbb{E}_\rho^{\mathbb{Q}^f} [\sigma_1\sigma_2/b \mid S_t^1/S_t^2]} - a(t, S_t^1, S_t^2) \right). \end{aligned} \quad (3.1)$$

We denote by $\rho_{(a,b)}$ a solution to (3.1).

PROOF Let $\rho \in \mathcal{C}$ be an admissible correlation. We can always pick two functions $a(t, S^1, S^2)$ and $b(t, S^1, S^2)$ such that b does not vanish and

$$a(t, S^1, S^2) + b(t, S^1, S^2)\rho(t, S^1, S^2) \equiv f\left(t, \frac{S^1}{S^2}\right)$$

¹ From now on, for the sake of clarity, we may omit the functions arguments t, S_t^1, S_t^2 , etc, within conditional expectations in long equations.

is local in cross. For instance, choose $b \equiv 1$ and $a(t, S^1, S^2) = f(t, S^1/S^2) - \rho(t, S^1, S^2)$ for some function f . Then,

$$\begin{aligned} \sigma_{12}^2 & \left(t, \frac{S_t^1}{S_t^2} \right) \\ &= \mathbb{E}_{\rho}^{\mathbb{Q}^f} \left[\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) - 2\rho(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2) \mid \frac{S_t^1}{S_t^2} \right] \\ &= \mathbb{E}_{\rho}^{\mathbb{Q}^f} \left[\sigma_1^2 + \sigma_2^2 + 2\frac{a}{b}\sigma_1\sigma_2 \mid \frac{S_t^1}{S_t^2} \right] - 2(a + b\rho) \left(t, \frac{S_t^1}{S_t^2} \right) \mathbb{E}_{\rho}^{\mathbb{Q}^f} \left[\frac{\sigma_1\sigma_2}{b} \mid \frac{S_t^1}{S_t^2} \right]. \end{aligned}$$

As a consequence, ρ satisfies (3.1). Conversely, if $\rho \in \mathcal{C}$ satisfies (3.1), then it satisfies (2.3) and is thus an admissible correlation. \square

We call (3.1) the local in cross affine transform representation of admissible correlations. The locality in cross of the affine transform $a + b\rho$ may seem an ad hoc assumption a priori, but it actually generates all the admissible correlations. One can try various pairs (a, b) and keep those pairs for which $\rho_{(a,b)} \in \mathcal{C}$. Among those, one can further choose the ones that fulfill some of the desirable properties \mathcal{P} .

The two already mentioned existing approaches to building admissible correlations correspond to two special cases of (a, b) . When $a \equiv 0$ and $b \equiv 1$, one assumes that the correlation itself is local in cross (see Guyon and Henry-Labordère 2011b; Reghai 2010). When $a = \sigma_1^2 + \sigma_2^2$ and $b = -2\sigma_1\sigma_2$, one assumes that the instantaneous variance of the cross-rate is local in cross (see Kovrizhkin 2012; Langnau 2010). We will come back to both examples and introduce new ones in Sections 5 and 9.

To the best of our knowledge, the existence and uniqueness of the nonlinear stochastic differential equations (SDEs) describing the calibrated models

$$\begin{aligned} dS_t^1 &= (r_t^d - r_t^1)S_t^1 dt + \sigma_1(t, S_t^1)S_t^1 dW_t^1, \\ dS_t^2 &= (r_t^d - r_t^2)S_t^2 dt + \sigma_2(t, S_t^2)S_t^2 dW_t^2, \\ d\langle W^1, W^2 \rangle_t &= \frac{dt}{b(t, S_t^1, S_t^2)} \\ &\times \left(\frac{\mathbb{E}^{\mathbb{Q}}[S_t^2(\sigma_1^2 + \sigma_2^2 + 2(a/b)\sigma_1\sigma_2) \mid S_t^1/S_t^2] - \sigma_{12}^2(t, S_t^1/S_t^2)\mathbb{E}^{\mathbb{Q}}[S_t^2 \mid S_t^1/S_t^2]}{2\mathbb{E}^{\mathbb{Q}}[S_t^2(\sigma_1\sigma_2/b) \mid S_t^1/S_t^2]} - a(t, S_t^1, S_t^2) \right) \end{aligned}$$

are still open mathematical questions. This is common to a variety of smile calibration problems (see, for instance, Guyon and Henry-Labordère (2011b, 2014) for an investigation of some nonlinear SDEs describing models calibrated to a smile as well as a discussion and numerical experiments on the existence of a solution). In practice, one can build a solution $\rho_{(a,b)} \in \mathcal{C}$ using the particle method of Guyon and Henry-Labordère (2011b), as we explain in the next section.

REMARK 3.3 As stated above, one can always require that $b \equiv 1$. Consequently, any correlation candidate is also of the subtype $\rho_{(a,1)}$:

$$\begin{aligned} & \rho_{(a,1)}(t, S_t^1, S_t^2) \\ &= \frac{1}{2\mathbb{E}_{\rho_{(a,1)}}^{\mathbb{Q}^f}[\sigma_1(t, S_t^1)\sigma_2(t, S_t^2) \mid S_t^1/S_t^2]} \\ & \quad \times \left(\mathbb{E}_{\rho_{(a,1)}}^{\mathbb{Q}^f} \left[\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) \right. \right. \\ & \quad \left. \left. + 2a(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2) \middle| \frac{S_t^1}{S_t^2} \right] - \sigma_{12}^2 \left(t, \frac{S_t^1}{S_t^2} \right) \right) \\ & \quad - a(t, S_t^1, S_t^2). \end{aligned} \tag{3.2}$$

The advantage of dealing with $a + b\rho$ instead of $a + \rho$ is that it includes the common approach in which $\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$ is assumed to be local in cross, with both $a = \sigma_1^2 + \sigma_2^2$ and $b = -2\sigma_1\sigma_2$ being independent of ρ . In general, one cannot require that $a \equiv 0$, because it would require that $\rho(t, S^1, S^2) = 0 \Rightarrow \rho(t, \lambda S^1, \lambda S^2) = 0$ for all $\lambda > 0$. Any admissible correlation satisfying the above condition – in particular, any nonvanishing admissible correlation – is also of the subtype $\rho_{(0,b)}$:

$$\rho_{(0,b)}(t, S_t^1, S_t^2) = \frac{\mathbb{E}_{\rho_{(0,b)}}^{\mathbb{Q}^f}[\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) \mid S_t^1/S_t^2] - \sigma_{12}^2(t, S_t^1/S_t^2)}{2b(t, S_t^1, S_t^2)\mathbb{E}_{\rho_{(0,b)}}^{\mathbb{Q}^f}[\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)/b(t, S_t^1, S_t^2) \mid S_t^1/S_t^2]}. \tag{3.3}$$

REMARK 3.4 Our affine transform technique is very robust. For instance, it allows us to consider more general models in which the instantaneous correlation depends on path-dependent variables as well. For instance, we can handle situations where ρ depends not only on (t, S_t^1, S_t^2) but also on the running averages $(1/t) \int_0^t S_u^1 du$ and $(1/t) \int_0^t S_u^2 du$, or

- on moving averages,
- on the running minimums and maximums of S^1 and S^2 ,
- on the realized correlation over the past few days,
- on the realized volatilities over the past few days, etc.

All one has to do is to also include the path-dependent variables in the arguments of the functions a and b . It is easy to adapt this remark to build single-asset path-dependent volatility models that calibrate to the smile (see Guyon 2014b). The calibration of cross-dependent volatility models, the multi-asset version of path-dependent volatility models, is discussed in Guyon (2016). The method also easily handles local stochastic volatility, stochastic interest rates and stochastic dividend yields, as explained in Section 10.

4 THE PARTICLE METHOD FOR LOCAL CORRELATION

The particle method for solving various smile calibration problems, including the calibration of local stochastic volatility models, with or without stochastic interest rates, and of local correlation models, was presented in Guyon and Henry-Labordère (2011b). It was also used in Jourdain and Sbai (2012) to calibrate a model coupling an index and its constituents. In the context presented in Section 3, the particle algorithm can be described as follows. Let $\{t_k\}$ denote a time discretization of $[0, T]$. We simulate N processes $(S_t^{1,i}, S_t^{2,i})_{1 \leq i \leq N}$, starting from (S_0^1, S_0^2) at time 0 and using N independent Brownian motions under the domestic measure \mathbb{Q} , as follows.

- (1) Initialize $k = 1$ and set

$$\rho_{(a,b)}(t, S^1, S^2) = \frac{\sigma_1^2(0, S^1) + \sigma_2^2(0, S^2) - \sigma_{12}^2(0, S^1/S^2)}{2\sigma_1(0, S^1)\sigma_2(0, S^2)}$$

for all $t \in [t_0 = 0; t_1]$. (At $t = 0$, no conditional expectation is computed, so $\rho_{(a,b)}$ does not depend on (a, b) .)

- (2) Simulate $(S_t^{1,i}, S_t^{2,i})_{1 \leq i \leq N}$ from t_{k-1} to t_k using a discretization scheme, say, a log-Euler scheme.
- (3) For all S^{12} in a grid G_{t_k} of cross-rate values, compute

$$\begin{aligned} E_{t_k}^{\text{num}}(S^{12}) &= \frac{1}{\sum_{i=1}^N S_{t_k}^{2,i} \delta_{t_k, N}((S_{t_k}^{1,i}/S_{t_k}^{2,i}) - S^{12})} \\ &\quad \times \left(\sum_{i=1}^N S_{t_k}^{2,i} (\sigma_1^2(t_k, S_{t_k}^{1,i}) + \sigma_2^2(t_k, S_{t_k}^{2,i}) + 2 \frac{a(t_k, S_{t_k}^{1,i}, S_{t_k}^{2,i})}{b(t_k, S_{t_k}^{1,i}, S_{t_k}^{2,i})} \right. \\ &\quad \left. \times \sigma_1(t_k, S_{t_k}^{1,i})\sigma_2(t_k, S_{t_k}^{2,i})) \delta_{t_k, N} \left(\frac{S_{t_k}^{1,i}}{S_{t_k}^{2,i}} - S^{12} \right) \right), \\ E_{t_k}^{\text{den}}(S^{12}) &= \frac{1}{\sum_{i=1}^N S_{t_k}^{2,i} \delta_{t_k, N}((S_{t_k}^{1,i}/S_{t_k}^{2,i}) - S^{12})} \\ &\quad \times \left(\sum_{i=1}^N S_{t_k}^{2,i} \frac{\sigma_1(t_k, S_{t_k}^{1,i})\sigma_2(t_k, S_{t_k}^{2,i})}{b(t_k, S_{t_k}^{1,i}, S_{t_k}^{2,i})} \delta_{t_k, N} \left(\frac{S_{t_k}^{1,i}}{S_{t_k}^{2,i}} - S^{12} \right) \right), \\ f(t_k, S^{12}) &= \frac{E_{t_k}^{\text{num}}(S^{12}) - \sigma_{12}^2(t_k, S^{12})}{2E_{t_k}^{\text{den}}(S^{12})}; \end{aligned}$$

interpolate and extrapolate $f(t_k, \cdot)$, for instance, using cubic splines; and, for all $t \in [t_k, t_{k+1}]$, set

$$\rho_{(a,b)}(t, S^1, S^2) = \frac{1}{b(t, S^1, S^2)} \left(f\left(t_k, \frac{S^1}{S^2}\right) - a(t, S^1, S^2) \right).$$

If $\rho_{(a,b)}(t, S^1, S^2) > 1$ (respectively < -1), then the trial (a, b) is a failure: $\rho_{(a,b)}$ is not admissible. Nonetheless, one can set $\rho_{(a,b)}(t, S^1, S^2) = 1$ (respectively -1) and carry on simulating S^1 and S^2 . The resulting calibration is then imperfect, but it may be accurate enough for trading purposes (see Figures 5, 6, 7, 8 and 10).

- (4) Set $k := k + 1$. Iterate steps (2) and (3) up to the maturity date T .

Here, $\delta_{t,N}(x) = (1/h_{t,N})K(x/h_{t,N})$ is an approximation of the Delta Dirac function; K is a fixed, symmetric, nonnegative kernel; and $h_{t,N}$ is a bandwidth that tends to zero as N grows to infinity. $E_t^{\text{num}}(S^{12})$ and $E_t^{\text{den}}(S^{12})$ approximate the conditional expectations

$$\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f} \left[\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) + 2 \frac{a(t, S_t^1, S_t^2)}{b(t, S_t^1, S_t^2)} \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) \mid \frac{S_t^1}{S_t^2} = S^{12} \right]$$

and

$$\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f} \left[\frac{\sigma_1(t, S_t^1) \sigma_2(t, S_t^2)}{b(t, S_t^1, S_t^2)} \mid \frac{S_t^1}{S_t^2} = S^{12} \right],$$

respectively. Alternative methods for estimating such conditional expectations include B-spline techniques, as explained in a recent work by Corlay (2013). Implementation details can be found in Section 9.

5 SOME EXAMPLES OF PAIRS OF FUNCTIONS (a, b)

As already mentioned in Section 3, the two existing approaches to building admissible correlations are special cases of the following local in cross affine transform representations.

- $a \equiv 0$ and $b \equiv 1$ (local in cross correlation model): in this case (Guyon and Henry-Labordère 2011b; Reghai 2010), one assumes that the correlation itself is local in cross:

$$\rho_{(0,1)}(t, S_t^1, S_t^2) = \frac{\mathbb{E}_{\rho_{(0,1)}}^{\mathbb{Q}^f} [\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) \mid S_t^1/S_t^2] - \sigma_{12}^2(t, S_t^1/S_t^2)}{2 \mathbb{E}_{\rho_{(0,1)}}^{\mathbb{Q}^f} [\sigma_1(t, S_t^1) \sigma_2(t, S_t^2) \mid S_t^1/S_t^2]}. \quad (5.1)$$

- $a = \sigma_1^2 + \sigma_2^2$ and $b = -2\sigma_1\sigma_2$ (local in cross volatility model): in this case, one assumes that the instantaneous variance of the cross-rate is local in cross. This is in the spirit of Langnau (2010) and has been studied in this FX context in Kovrizhkin (2012). We denote by ρ^* the correlation candidate

$$\rho^*(t, S_t^1, S_t^2) = \frac{\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) - \sigma_{12}^2(t, S_t^1/S_t^2)}{2\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)}. \quad (5.2)$$

This is the only situation in which no estimation of conditional expectation (given the value of S_t^1/S_t^2) is needed. As a consequence, ρ^* is well defined, even if it exits the interval $[-1, 1]$.

Another natural choice of (a, b) is the following.

- $a \equiv 0$ and $b = \sigma_1\sigma_2$ (local in cross covariance model): in this case, one assumes that the local covariance $\rho(t, S^1, S^2)\sigma_1(t, S^1)\sigma_2(t, S^2)$ of increments of S^1 and S^2 is local in cross. This choice defines a model calibrated to the three FX smiles if and only if

$$\frac{\mathbb{E}_{\rho(0,\sigma_1\sigma_2)}^{\mathbb{Q}^f} [\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) | S_t^1/S_t^2] - \sigma_{12}^2(t, S_t^1/S_t^2)}{2\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)} \in [-1, 1].$$

REMARK 5.1 In general, the instantaneous volatility of the cross

$$\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

in the model depends jointly on S^1 and S^2 , and not only on S^{12} . The instantaneous volatility of S^1 (respectively S^2), however, depends only on S^1 (respectively S^2). This means that, in general, the model is not symmetric in the three FX rates. If ρ^* takes values in $[-1, 1]$, then the local in cross volatility model is the unique local volatility–local correlation model that is symmetric in the three FX rates. Otherwise, no such model exists. The asymmetry may not necessarily be seen as a drawback: in cases where a currency, eg, USD, is “stronger” than the other two, eg, BRL and TRY, it is natural to choose it as the anchor rate, and it might make sense to price and hedge in a model where the volatility of S^{12} (BRL/TRY) is actually a function of S^1 (BRL/USD) and S^2 (TRY/USD) separately, rather than a function of S^{12} only. Note, however, that if one changes the anchor rate, one gets another family of admissible local correlation models, and different prices for exotic options (see, for example, De Col *et al* (2013) for a symmetric way to approximately calibrate to a triangle of FX market smiles).

6 SOME LINKS BETWEEN LOCAL CORRELATIONS

Assume that $\rho_{(a,b)}$ is an admissible correlation. We can express the affine transform $a + b\rho_{(a,b)}$ of $\rho_{(a,b)}$ as an average of the same affine transform of the correlation candidate ρ^* (even if ρ^* takes values outside $[-1, 1]$):

$$\begin{aligned} & (a + b\rho_{(a,b)})(t, \frac{S_t^1}{S_t^2}) \\ &= \frac{\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f}[(a + b\rho^*)(t, S_t^1, S_t^2)(\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)/b(t, S_t^1, S_t^2)) \mid S_t^1/S_t^2]}{\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f}[\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)/b(t, S_t^1, S_t^2) \mid S_t^1/S_t^2]} \\ &\equiv \mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^{\sigma_1\sigma_2/b}} \left[(a + b\rho^*)(t, S_t^1, S_t^2) \mid \frac{S_t^1}{S_t^2} \right], \end{aligned}$$

where

$$\frac{d\mathbb{Q}^{\sigma_1\sigma_2/b}}{d\mathbb{Q}^f} \equiv \frac{\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)/b(t, S_t^1, S_t^2)}{\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f}[\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)/b(t, S_t^1, S_t^2)]}.$$

In particular, if $\rho_{(0,1)}$ is an admissible correlation, $\rho_{(0,1)}$ is a weighted average of ρ^* on each “constant cross” line:

$$\begin{aligned} \rho_{(0,1)}\left(t, \frac{S_t^1}{S_t^2}\right) &= \frac{\mathbb{E}_{\rho_{(0,1)}}^{\mathbb{Q}^f}[\rho^*(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2) \mid S_t^1/S_t^2]}{\mathbb{E}_{\rho_{(0,1)}}^{\mathbb{Q}^f}[\sigma_1(t, S_t^1)\sigma_2(t, S_t^2) \mid S_t^1/S_t^2]} \\ &\equiv \mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^{\sigma_1\sigma_2}} \left[\rho^*(t, S_t^1, S_t^2) \mid \frac{S_t^1}{S_t^2} \right]. \end{aligned} \quad (6.1)$$

This has two consequences.

- If $\rho_{(0,1)}$ is an admissible correlation, then its image, ie, the range of values it takes, is included in the image of ρ^* .
- $\tau_{\rho_{(0,1)}} \geq \tau_{\rho^*}$, where τ_ρ denotes the smallest time at which ρ fails to be a correlation function:

$$\tau_\rho = \inf\{t \in [0, T] \mid \exists S^1, S^2 > 0, \rho(t, S^1, S^2) \notin [-1, 1]\}. \quad (6.2)$$

In the particular case where σ_1 and σ_2 depend only on t (no skew on S^1 or S^2), then (2.3) simply reads

$$\sigma_{12}^2\left(t, \frac{S_t^1}{S_t^2}\right) = \sigma_1(t)^2 + \sigma_2(t)^2 - 2\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f} \left[\rho_{(a,b)}(t, S_t^1, S_t^2) \mid \frac{S_t^1}{S_t^2} \right] \sigma_1(t)\sigma_2(t),$$

ie, using (5.1) and (5.2),

$$\rho^*(t, S_t^1, S_t^2) = \rho_{(0,1)}\left(t, \frac{S_t^1}{S_t^2}\right) = \mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f} \left[\rho_{(a,b)}(t, S_t^1, S_t^2) \mid \frac{S_t^1}{S_t^2} \right]. \quad (6.3)$$

Note that, in this case, all the examples of Section 5 boil down to the same correlation model, in which the correlation is local in cross. In this situation, $\rho_{(0,1)} = \rho^*$ is, among all the admissible correlations $\rho(t, S^1, S^2)$, the one with the smallest image. This means that if $\rho_{(0,1)} = \rho^*$ is not admissible, then no correlation $\rho(t, S^1, S^2)$ is admissible:

$$\rho_{(0,1)}\left(t, \frac{S_t^1}{S_t^2}\right) = \rho^*\left(t, \frac{S_t^1}{S_t^2}\right) > 1$$

corresponds to the situation where

$$|\sigma_1(t) - \sigma_2(t)| > \sigma_{12}\left(t, \frac{S_t^1}{S_t^2}\right),$$

while

$$\rho_{(0,1)}\left(t, \frac{S_t^1}{S_t^2}\right) = \rho^*\left(t, \frac{S_t^1}{S_t^2}\right) < -1$$

corresponds to the situation where

$$\sigma_1(t) + \sigma_2(t) < \sigma_{12}\left(t, \frac{S_t^1}{S_t^2}\right).$$

Equation (6.3) tells us that in the case where S^1 and S^2 have no skew, all admissible correlations have the same average value under \mathbb{Q}^f over each constant cross line, and this common average value is $\rho_{(0,1)}$.

7 PRICE IMPACT OF CORRELATION

Different choices of admissible $\rho_{(a,b)}$ will lead to different prices for exotic options on S^1 and S^2 , while still producing the same prices for vanilla options on S^1 , S^2 and the cross-rate S^1/S^2 . In this section, we analyze the impact of $\rho_{(a,b)}$ on the price of options on S^1 and S^2 . This helps us to develop an intuition of the model and choose an adequate model for pricing and hedging a given option.

7.1 The price impact formula

Here, we follow a reasoning inspired by El Karoui *et al* (1998) and Dupire (1998) to quantify the impact of the correlation model on the price of the option that delivers the payout $g(S_T^1, S_T^2)$ in domestic currency (USD in our example) at maturity T .

PROPOSITION 7.1 *Let $\rho_0(t, S^1, S^2)$ be a correlation function, and let $P_0(t, S^1, S^2)$ be the corresponding pricing function in domestic currency (model \mathcal{M}_{ρ_0}). We consider a process (S_t^1, S_t^2) whose dynamics derive not from ρ_0 but from a general correlation process ρ_t (model \mathcal{M}_{ρ_t}). Then, under integrability assumptions, the price difference*

between model \mathcal{M}_{ρ_t} and model \mathcal{M}_{ρ_0} is the expected value of the integrated discounted tracking error

$$\begin{aligned} D_{0T}^d \mathbb{E}_{\rho_t}^{\mathbb{Q}}[g(S_T^1, S_T^2)] - P_0(0, S_0^1, S_0^2) \\ = \mathbb{E}_{\rho_t}^{\mathbb{Q}} \left[\int_0^T D_{0t}^d (\rho_t - \rho_0(t, S_t^1, S_t^2)) \right. \\ \times \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_0(t, S_t^1, S_t^2) dt \left. \right]. \quad (7.1) \end{aligned}$$

We use the notation $\mathbb{E}_{\rho_t}^{\mathbb{Q}}$ to emphasize that the process (S_t^1, S_t^2) is simulated under model \mathcal{M}_{ρ_t} . The instantaneous tracking error at date t ,

$$\epsilon_t \equiv (\rho_t - \rho_0(t, S_t^1, S_t^2)) \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_0(t, S_t^1, S_t^2),$$

consists of the spread of the two correlations times the \mathcal{M}_{ρ_0} -cross Gamma $\partial_{S^1 S^2}^2 P_0$, times the product of (normal) volatilities $\sigma_1 \sigma_2 S^1 S^2$, where all terms are evaluated at the spots (S_t^1, S_t^2) defined by model \mathcal{M}_{ρ_t} . The interpretation as an error comes from the fact that $\epsilon_t dt$ is the infinitesimal P&L between t and $t + dt$ of a Delta-hedged long position in one option, when one uses the pricing function P_0 and the corresponding Deltas $\partial_{S^1} P_0, \partial_{S^2} P_0$, derived from model \mathcal{M}_{ρ_0} , while actual dynamics of the assets is given by model \mathcal{M}_{ρ_t} .

PROOF P_0 is the solution to the backward partial differential equation (PDE) $(\partial_t + \mathcal{L})P_0 = 0$, $P_0(T, S^1, S^2) = g(S^1, S^2)$, where

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sigma_1^2(t, S^1) (S^1)^2 \partial_{S^1}^2 + \frac{1}{2} \sigma_2^2(t, S^2) (S^2)^2 \partial_{S^2}^2 \\ & + \rho_0(t, S^1, S^2) \sigma_1(t, S^1) \sigma_2(t, S^2) S^1 S^2 \partial_{S^1 S^2}^2 \\ & + (r_t^d - r_t^1) S^1 \partial_{S^1} + (r_t^d - r_t^2) S^2 \partial_{S^2} - r_t^d. \end{aligned}$$

Applying Ito's formula to the function P_0 and the process (S_t^1, S_t^2) , and using the above PDE, we get

$$\begin{aligned} D_{0T}^d g(S_T^1, S_T^2) - P_0(0, S_0^1, S_0^2) \\ = \int_0^T D_{0t}^d (\partial_t P_0(t, S_t^1, S_t^2) + \frac{1}{2} \sigma_1^2(t, S_t^1) (S_t^1)^2 \partial_{S^1}^2 P_0(t, S_t^1, S_t^2) \\ + \frac{1}{2} \sigma_2^2(t, S_t^2) (S_t^2)^2 \partial_{S^2}^2 P_0(t, S_t^1, S_t^2) \\ + \rho_t \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_0(t, S_t^1, S_t^2) \\ + (r_t^d - r_t^1) S_t^1 \partial_{S^1} P_0(t, S_t^1, S_t^2) \\ + (r_t^d - r_t^2) S_t^2 \partial_{S^2} P_0(t, S_t^1, S_t^2) - r_t^d P_0(t, S_t^1, S_t^2)) dt + M_t \end{aligned}$$

$$= \int_0^T D_{0t}^d (\rho_t - \rho_0(t, S_t^1, S_t^2)) \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_0(t, S_t^1, S_t^2) dt \\ + M_t,$$

where $D_{0T}^d = \exp(-\int_0^T r_t^d dt)$ is the (deterministic) discount factor, and

$$M_t = \int_0^T D_{0t}^d \partial_{S^1} P_0(t, S_t^1, S_t^2) \sigma_1(t, S_t^1) S_t^1 dW_t^1 \\ + \int_0^T D_{0t}^d \partial_{S^2} P_0(t, S_t^1, S_t^2) \sigma_2(t, S_t^2) S_t^2 dW_t^2$$

is a local martingale under the risk-neutral measure \mathbb{Q} . Under integrability conditions, it is a true martingale and, taking expectations, we get (7.1). \square

Equation (7.1) has several interesting consequences that we address in Sections 7.2, 7.3 and 7.4.

7.2 Equivalent local correlation

From (7.1), by conditioning on (S_t^1, S_t^2) , we get

$$D_{0T}^d \mathbb{E}_{\rho_t}^{\mathbb{Q}} [g(S_T^1, S_T^2)] - P_0(0, S_0^1, S_0^2) \\ = \mathbb{E}_{\rho_t}^{\mathbb{Q}} \left[\int_0^T D_{0t}^d (\rho_{\text{loc}}(t, S_t^1, S_t^2) - \rho_0(t, S_t^1, S_t^2)) \right. \\ \times \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_0(t, S_t^1, S_t^2) dt \left. \right], \quad (7.2)$$

where $\rho_{\text{loc}}(t, S_t^1, S_t^2) \equiv \mathbb{E}_{\rho_t}^{\mathbb{Q}} [\rho_t \mid S_t^1, S_t^2]$ is the equivalent local correlation. From Gyöngy's theorem (1986), we know that the model, say $\mathcal{M}_{\rho_{\text{loc}}}$, that uses the local correlation function ρ_{loc} generates the same distributions for (S_t^1, S_t^2) as model \mathcal{M}_{ρ_t} , for all t . This can be easily re-derived by applying (7.2), with model $\mathcal{M}_{\rho_{\text{loc}}}$ playing the role of \mathcal{M}_{ρ_0} :

$$D_{0T}^d \mathbb{E}_{\rho_t}^{\mathbb{Q}} [g(S_T^1, S_T^2)] - P_{\rho_{\text{loc}}}(0, S_0^1, S_0^2) \\ = \mathbb{E}_{\rho_t}^{\mathbb{Q}} \left[\int_0^T D_{0t}^d (\rho_{\text{loc}}(t, S_t^1, S_t^2) - \rho_{\text{loc}}(t, S_t^1, S_t^2)) \right. \\ \times \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho_{\text{loc}}}(t, S_t^1, S_t^2) dt \left. \right] = 0.$$

This proves that all vanilla payoffs $g(S_T^1, S_T^2)$ have identical prices in models \mathcal{M}_{ρ_t} and $\mathcal{M}_{\rho_{\text{loc}}}$, ie, (S_T^1, S_T^2) have identical distributions under \mathbb{Q} in both models.

7.3 Implied correlation

Equation (7.1), or equivalently (7.2), also allows us to define the implied correlation. Given a general model \mathcal{M}_{ρ_t} and a payoff $g(S_T^1, S_T^2)$, we define the implied correlation $\rho(T, g)$ as the value of the constant correlation, such that the option has the same price in model \mathcal{M}_{ρ_t} and the model with constant correlation, ie, such that

$$\mathbb{E}_{\rho_t}^{\mathbb{Q}} \left[\int_0^T D_{0t}^d (\rho_t - \rho(T, g)) \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(T,g)}(t, S_t^1, S_t^2) dt \right] = 0,$$

or equivalently

$$\rho(T, g) = \frac{\mathbb{E}_{\rho_t}^{\mathbb{Q}} [\int_0^T \rho_t D_{0t}^d \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(T,g)}(t, S_t^1, S_t^2) dt]}{\mathbb{E}_{\rho_t}^{\mathbb{Q}} [\int_0^T D_{0t}^d \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(T,g)}(t, S_t^1, S_t^2) dt]}. \quad (7.3)$$

This is similar to Dupire's expression of implied volatility as a weighted average of the spot volatility (Dupire 1998). Note that (7.3) is a fixed point equation, because the right-hand side depends on $\rho(T, g)$ as well through the cross Gamma $\partial_{S^1 S^2}^2 P_{\rho(T,g)}$. If (7.3) admits a unique solution, then the implied correlation exists and is uniquely defined.

Following Guyon and Henry-Labordère (2011a), one can estimate the implied correlation by estimating the fixed point of the mapping

$$\rho \mapsto \int_0^T \int_0^\infty \int_0^\infty \rho_{loc}(t, S^1, S^2) q_\rho(t, S^1, S^2) dS^1 dS^2 dt,$$

where

$$\begin{aligned} q_\rho(t, S^1, S^2) &= D_{0t}^d \sigma_1(t, S^1) \sigma_2(t, S^2) S^1 S^2 \partial_{S^1 S^2}^2 P_\rho(t, S^1, S^2) p(t, S^1, S^2) \\ &\times \left[\int_0^T \int_0^\infty \int_0^\infty D_{0t_*}^d \sigma_1(t_*, S_*^1) \sigma_2(t_*, S_*^2) S_*^1 S_*^2 \partial_{S^1 S^2}^2 \right. \\ &\quad \left. \times P_\rho(t_*, S_*^1, S_*^2) p(t_*, S_*^1, S_*^2) dS_*^1 dS_*^2 dt_* \right]^{-1}, \end{aligned}$$

with $p(t, S^1, S^2)$ the probability density function of (S_t^1, S_t^2) when the correlation is ρ_t , or, equivalently, $\rho_{loc}(t, S_t^1, S_t^2)$. Of course, the density $p(t, S^1, S^2)$ is unknown; otherwise, we could compute exactly the price of the option. One way to estimate the implied correlation is to compute the fixed point of the approximate mapping, where $p(t, S^1, S^2)$ is replaced by some explicit estimate $\hat{p}(t, S^1, S^2)$. In the particular

case where the instantaneous volatilities and correlation are constant, the weight $q_\rho(t, S^1, S^2)$ is known explicitly for the payoff $g(S_T^1, S_T^2) = (S_T^1 - KS_T^2)_+$. Figure 1 shows the graphs of $(S^1, S^2) \mapsto q_\rho(t, S^1, S^2)$ for increasing values of t , from 0 to T .

Following Gatheral (2006) (see also Guyon and Henry-Labordère 2011a), we can get an alternative expression for the implied correlation by considering the situation where the local correlation function $\rho(t)$ is a deterministic function of time only. The option has the same price in model \mathcal{M}_{ρ_t} and in this model if and only if

$$\int_0^T D_{0t}^d \mathbb{E}_{\rho_t}^\mathbb{Q} [(\rho_t - \rho(t)) \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t)}(t, S_t^1, S_t^2)] dt = 0.$$

There is a unique function $\rho(t) \equiv \rho(t; T, g)$, such that, not only is the time integral zero, but also the integrand vanishes for each time slice t :

$$\rho(t; T, g) = \frac{\mathbb{E}_{\rho_t}^\mathbb{Q} [\rho_t \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t; T, g)}(t, S_t^1, S_t^2)]}{\mathbb{E}_{\rho_t}^\mathbb{Q} [\sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t; T, g)}(t, S_t^1, S_t^2)]}.$$

Note that this is again a fixed point equation, because the right-hand side depends on $\rho(t; T, g)$ through the cross Gamma $\partial_{S^1 S^2}^2 P_{\rho(t; T, g)}$. In the particular case where $\sigma_1(t, S_t^1) = \sigma_1(t)$ and $\sigma_2(t, S_t^2) = \sigma_2(t)$ depend only on time (no volatility skew on S^1 and S^2), then

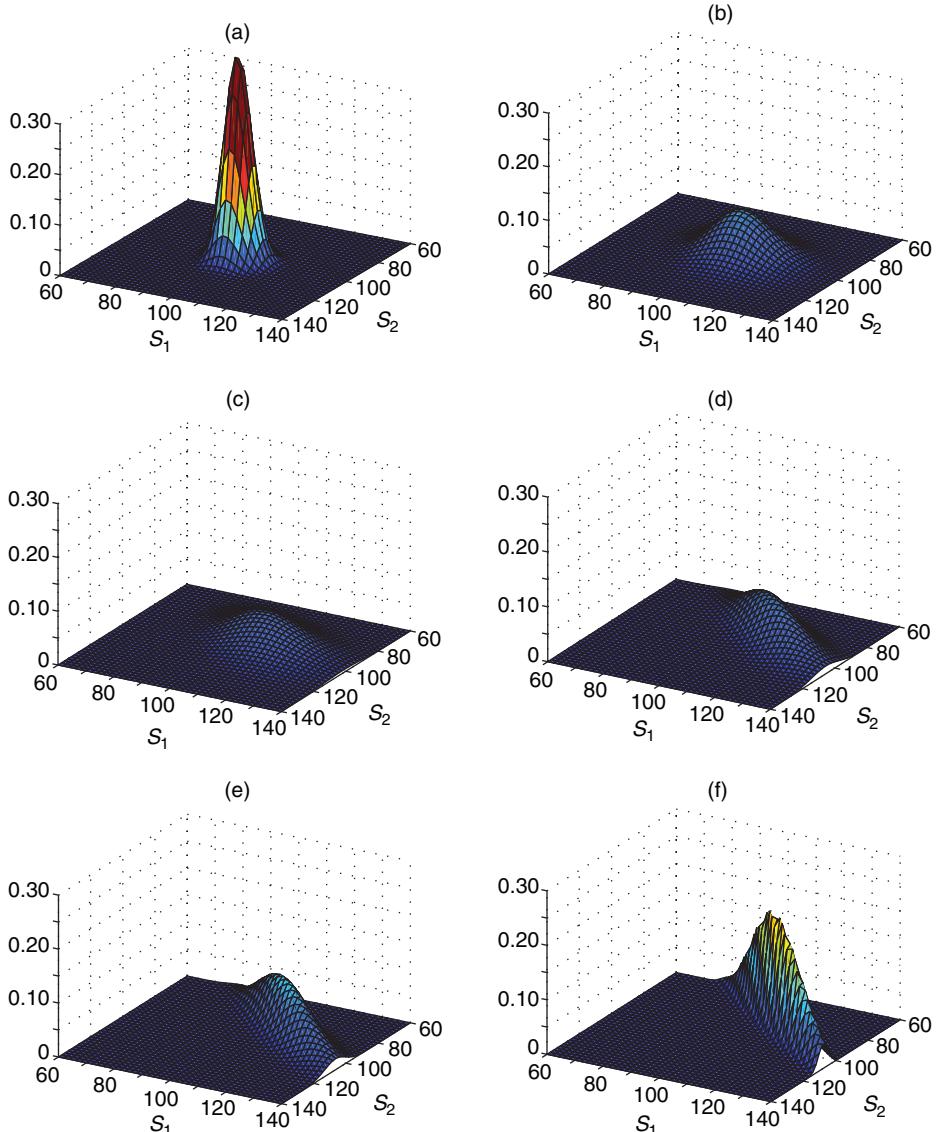
$$\begin{aligned} \rho(T, g) &= \frac{\int_0^T \rho(t; T, g) \sigma_1(t) \sigma_2(t) dt}{\int_0^T \sigma_1(t) \sigma_2(t) dt} \\ &= \frac{1}{\int_0^T \sigma_1(t) \sigma_2(t) dt} \\ &\quad \times \left(\int_0^T \frac{\mathbb{E}_{\rho_t}^\mathbb{Q} [\rho_t S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t; T, g)}(t, S_t^1, S_t^2)]}{\mathbb{E}_{\rho_t}^\mathbb{Q} [S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t; T, g)}(t, S_t^1, S_t^2)]} \sigma_1(t) \sigma_2(t) dt \right). \end{aligned} \tag{7.4}$$

In the even more particular case where σ_1 and σ_2 are constant, this reads

$$\rho(T, g) = \frac{1}{T} \int_0^T \frac{\mathbb{E}_{\rho_t}^\mathbb{Q} [\rho_t S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t; T, g)}(t, S_t^1, S_t^2)]}{\mathbb{E}_{\rho_t}^\mathbb{Q} [S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_{\rho(t; T, g)}(t, S_t^1, S_t^2)]} dt. \tag{7.5}$$

Equations (7.4) and (7.5) are similar to Gatheral's formula for the implied volatility (see Gatheral 2006; Guyon and Henry-Labordère 2011a). Note, however, that we had to assume no skew on S^1 and S^2 to derive them. When S^1 or S^2 have a skew, we can

FIGURE 1 Graphs of $(S^1, S^2) \mapsto q_\rho(t, S^1, S^2)$ for increasing values of t ((a) 0.05, (b) 0.30, (c) 0.60, (d) 0.90, (e) 0.95, (f) 0.99).



Black–Scholes model: $\sigma_1 = 20\%$, $\sigma_2 = 20\%$, $\rho = 0$, $S_0^1 = 100$, $S_0^2 = 100$. Payoff $g(S_T^1, S_T^2) = (S_T^1 - KS_T^2)_+$, $K = 1.3$, $T = 1$.

no longer easily link the implied volatility $\rho(T, g)$ to $\rho(t; T, g)$. Equation (7.5) looks like (7.3), but it actually differs in two ways:

- (i) (7.5) involves a time average of a space average, whereas (7.3) involves a joint time and space average;
- (ii) the cross Gammas involved in both equations differ slightly: (7.5) uses the cross Gamma computed with time-dependent correlation $\rho(t; T, g)$, whereas (7.3) uses the cross Gamma computed with constant correlation $\rho(T, g)$.

7.4 Impact of correlation on price

For a given option, (7.1) helps us to understand the impact on the option price of a particular choice of local correlation function. Basically, high option prices correspond to local correlation functions that are large in the region where the cross Gamma is positive, and small (ie, close to -1) in the region where the cross Gamma is negative. Of course, the cross Gamma depends on the particular model picked. Equation (7.1) states that to compute the price difference between model \mathcal{M}_{ρ_t} and reference model \mathcal{M}_{ρ_0} , you may compute the right-hand side expectation, where (S_t^1, S_t^2) is simulated under model \mathcal{M}_{ρ_t} and the cross Gamma is computed under model \mathcal{M}_{ρ_0} . Actually, in the case where the process ρ_t is a local correlation process $\rho_1(t, S_t^1, S_t^2)$, the roles of \mathcal{M}_{ρ_0} and \mathcal{M}_{ρ_1} can be swapped:

$$\begin{aligned} D_{0T}^d \mathbb{E}_{\rho_0}^{\mathbb{Q}}[g(S_T^1, S_T^2)] - P_1(0, S_0^1, S_0^2) \\ = \mathbb{E}_{\rho_0}^{\mathbb{Q}} \left[\int_0^T D_{0t}^d(\rho_0(t, S_t^1, S_t^2) - \rho_1(t, S_t^1, S_t^2)) \right. \\ \times \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_1(t, S_t^1, S_t^2) dt \left. \right], \end{aligned}$$

or equivalently

$$\begin{aligned} P_1(0, S_0^1, S_0^2) - P_0(0, S_0^1, S_0^2) \\ = \mathbb{E}_{\rho_0}^{\mathbb{Q}} \left[\int_0^T D_{0t}^d(\rho_1(t, S_t^1, S_t^2) - \rho_0(t, S_t^1, S_t^2)) \right. \\ \times \sigma_1(t, S_t^1) \sigma_2(t, S_t^2) S_t^1 S_t^2 \partial_{S^1 S^2}^2 P_1(t, S_t^1, S_t^2) dt \left. \right]. \end{aligned}$$

Stated otherwise, to compute the price difference between two local correlation models \mathcal{M}_{ρ_1} and \mathcal{M}_{ρ_0} , you may compute the expected value of the integrated tracking error. Here,

- either (S_t^1, S_t^2) is simulated under model \mathcal{M}_{ρ_1} , and the cross Gamma is computed under model \mathcal{M}_{ρ_0} ,
- or (S_t^1, S_t^2) is simulated under model \mathcal{M}_{ρ_0} , and the cross Gamma is computed under model \mathcal{M}_{ρ_1} .

7.5 Uncertain correlation model

The highest possible price for payoff g , given local volatility dynamics for S^1 and S^2 , is

$$D_{0T}^d \sup_{\rho_t \in \mathcal{R}} \mathbb{E}_{\rho_t}^{\mathbb{Q}} [g(S_T^1, S_T^2)],$$

where \mathcal{R} denotes the set of all adapted stochastic processes taking values in $[-1, 1]$. It is given by the solution $P(0, S_0^1, S_0^2)$ to the (nonlinear) Hamilton–Jacobi–Bellman (HJB) equation:

$$\begin{aligned} \partial_t P + \frac{1}{2}\sigma_1^2(t, S^1)(S^1)^2\partial_{S^1}^2 P + \frac{1}{2}\sigma_2^2(t, S^2)(S^2)^2\partial_{S^2}^2 P \\ + \sup_{\rho \in [-1, 1]} \{\rho\sigma_1(t, S^1)\sigma_2(t, S^2)S^1S^2\partial_{S^1S^2}^2 P\} \\ + (r_t^d - r_t^1)S^1\partial_{S^1} P + (r_t^d - r_t^2)S^2\partial_{S^2} P - r_t^d P = 0, \end{aligned}$$

with terminal condition $P(T, S^1, S^2) = g(S^1, S^2)$; that is,

$$\begin{aligned} \partial_t P + \frac{1}{2}\sigma_1^2(t, S^1)(S^1)^2\partial_{S^1}^2 P + \frac{1}{2}\sigma_2^2(t, S^2)(S^2)^2\partial_{S^2}^2 P \\ + \rho(\partial_{S^1S^2}^2 P)\sigma_1(t, S^1)\sigma_2(t, S^2)S^1S^2\partial_{S^1S^2}^2 P \\ + (r_t^d - r_t^1)S^1\partial_{S^1} P + (r_t^d - r_t^2)S^2\partial_{S^2} P - r_t^d P = 0, \quad (7.6) \end{aligned}$$

where $+1\rho(\Gamma) = 1$ if $\Gamma \geq 0$, -1 otherwise. As expected from (7.1), the highest option price corresponds to the local correlation function that is worth $+1$ in the region where the cross Gamma is positive, and -1 in the region where the cross Gamma is negative. Here, for consistency, the cross Gamma must be computed within this extremal model \mathcal{M}_{HJB} , ie, by solving (7.6). Symmetrically, the lower bound $D_{0T}^d \inf_{\rho_t \in \mathcal{R}} \mathbb{E}_{\rho_t} [g(S_T^1, S_T^2)]$ is given by $P(0, S_0^1, S_0^2)$, where P is the solution to (7.6), where $+\rho(\partial_{S^1S^2}^2 P)$ is replaced by $-\rho(\partial_{S^1S^2}^2 P)$.

8 THE EQUITY INDEX SMILE CALIBRATION PROBLEM

Let us now see how the reasoning presented above in Sections 2 and 3 for the FX smile triangle calibration problem can be extended to the N -dimensional equity index smile calibration problem. Let us consider an index $I_t = \sum_{i=1}^N \alpha_i S_t^i$ made of N weighted stocks, each of which is modeled using its own local volatility:

$$dS_t^i = r_t S_t^i dt + \sigma_i(t, S_t^i) S_t^i dW_t^i, \quad d\langle W^i, W^j \rangle_t = \rho_{ij}(t, S_t) dt. \quad (8.1)$$

The interest rate r_t is deterministic; $\{W^i\}$ denotes a multidimensional Brownian motion with an instantaneous correlation function of the time and the N stock values $S_t = (S_t^1, \dots, S_t^N)$. To simplify our notation, let us denote by

$$v_{\rho}(t, S_t) = \sum_{i,j=1}^N \alpha_i \alpha_j \rho_{ij}(t, S_t) \sigma_i(t, S_t^i) \sigma_j(t, S_t^j) S_t^i S_t^j$$

the instantaneous (normal) variance of the basket of stocks within model (8.1). From Avellaneda *et al* (2002), we know that the model is calibrated to the index smile if and only if

$$I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 = \mathbb{E}_\rho[v_\rho(t, S_t) | I_t], \quad (8.2)$$

where σ_{Dup}^I denotes the Dupire local volatility of the index (see proof in the online appendix).

REMARK 8.1 It is tempting but wrong to believe that in order to exclude arbitrage opportunities we must have the stronger statement $I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 = v_\rho(t, S_t)$ at each point in time. This is incorrect in two ways. First, only the conditional expected value of the basket variance given the index value matters. Second, if no ρ satisfies (8.2), it does not mean that arbitrage opportunities exist; rather, it means that prices are inconsistent with local volatilities–local correlation modeling, and that one has to consider more general models, for instance, models that include cross-dependent volatility (Guyon 2015) or stochastic volatility.

DEFINITION 8.2 Let \mathcal{C} denote the set of functions $\rho(t, S)$ taking values in the set of correlation matrixes. Any function $\rho \in \mathcal{C}$ satisfying (8.2) is called an admissible correlation.

We aim at characterizing and building the family of all admissible correlations. Similar to Proposition 2.1, we have the following.

PROPOSITION 8.3 *Let $\rho \in \mathcal{C}$. It is an admissible correlation if and only if there exist two functions $\rho^0, \rho^1 \in \mathcal{C}$ and two real-valued functions a and b such that b does not vanish and*

$$\begin{aligned} \rho(t, S) &\equiv (1 - \lambda(t, S))\rho^0(t, S) + \lambda(t, S)\rho^1(t, S) \in \mathcal{C}, \\ \lambda(t, S_t) &= \frac{1}{b(t, S_t)} \\ &\times \left(\frac{1}{\mathbb{E}_\rho[(1/b(t, S_t))(v_{\rho^1}(t, S_t) - v_{\rho^0}(t, S_t)) | I_t]} \right. \\ &\quad \times \left(I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 - \mathbb{E}_\rho[v_{\rho^0}(t, S_t) \right. \\ &\quad \left. \left. - \frac{a(t, S_t)}{b(t, S_t)}(v_{\rho^1}(t, S_t) - v_{\rho^0}(t, S_t)) | I_t] \right) \right) \\ &\quad - a(t, S_t). \end{aligned} \quad (8.3)$$

We denote by $\lambda_{(\rho^0, \rho^1, a, b)}$ a solution to (8.3) and write $\rho_{(\rho^0, \rho^1, a, b)} \equiv (1 - \lambda_{(\rho^0, \rho^1, a, b)})\rho^0 + \lambda_{(\rho^0, \rho^1, a, b)}\rho^1$.

PROOF Assume that $\rho \in \mathcal{C}$ is admissible. Pick $\rho^0, \rho^1 \in \mathcal{C}$ such that ρ lies on the line defined by ρ^0 and ρ^1 :

$$\rho(t, S) = (1 - \lambda(t, S))\rho^0(t, S) + \lambda(t, S)\rho^1(t, S).$$

We can always pick two functions a and b such that b does not vanish and $a(t, S_t) + b(t, S_t)\lambda(t, S_t)$ is local in index, ie, is a function of (t, I_t) only, say $f(t, I_t)$. For instance, choose $b \equiv 1$ and $a(t, S_t) = f(t, I_t) - \lambda(t, S_t)$ for some function f . Then, (8.2) reads

$$\begin{aligned} I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 &= \mathbb{E}_\rho[v_{\rho^0}(t, S_t) + (v_{\rho^1}(t, S_t) - v_{\rho^0}(t, S_t))\lambda(t, S_t) \mid I_t] \\ &= (a + b\lambda)(t, I_t) \mathbb{E}_\rho \left[\frac{v_{\rho^1} - v_{\rho^0}}{b} \mid I_t \right] \\ &\quad + \mathbb{E}_\rho \left[v_{\rho^0} - \frac{a}{b}(v_{\rho^1} - v_{\rho^0}) \mid I_t \right], \end{aligned}$$

and $\lambda = \lambda_{(\rho^0, \rho^1, a, b)}$ satisfies (8.3). Conversely, if there exist two functions $\rho^0, \rho^1 \in \mathcal{C}$ and two real-valued functions a and b such that b does not vanish and (8.3) holds, then ρ satisfies (8.2) and is an admissible correlation. \square

For a given quadruplet $q \equiv (\rho^0, \rho^1, a, b)$, ρ_q is guaranteed to be PSD if λ_q takes values in $[0, 1]$, because the set of correlation matrixes is convex. When ρ^0 (respectively ρ^1) does not belong to the boundary of the set of correlation matrixes, ρ_q may be a correlation matrix even if $\lambda_q < 0$ (respectively $\lambda_q > 1$). We call the resulting model the local in index $a + b\lambda$ model. It also depends on the choice of ρ^0 and ρ^1 .

Equation (8.3) provides a constructive way to build all the admissible correlations using the particle method. One tries all the different quadruplets $q \equiv (\rho^0, \rho^1, a, b)$ and builds $\rho_q(t, S)$. If at some date $t < T$ and for some S , $\rho_q(t, S)$ is not PSD, then the trial is a failure: ρ_q is not admissible. Nevertheless, one may project ρ_q on the set of correlation matrixes, carry on the simulation, and decide if the (imperfect) calibration is accurate enough. This will be the case when $\rho_q(t, S)$ fails to be PSD only for unlikely values of S , under the risk-neutral distribution.

The two existing approaches correspond to special cases of this formulation.

- $a \equiv 0$ and $b \equiv 1$ (local in index λ model (Guyon and Henry-Labordère 2011b)): in this case, one assumes that the linear combination parameter λ itself is local in index:

$$\lambda_{(\rho^0, \rho^1, 0, 1)}(t, S_t) = \frac{I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 - \mathbb{E}_{\rho_{(\rho^0, \rho^1, 0, 1)}}[v_{\rho^0}(t, S_t) \mid I_t]}{\mathbb{E}_{\rho_{(\rho^0, \rho^1, 0, 1)}}[v_{\rho^1}(t, S_t) - v_{\rho^0}(t, S_t) \mid I_t]}.$$

In Reghai (2010), this same model is investigated, but no explicit formula is given for $\lambda_{(\rho^0, \rho^1, 0, 1)}$; instead, $\lambda_{(\rho^0, \rho^1, 0, 1)}$ is computed as the fixed point of a

mapping that requires computing basket implied volatilities at each iteration, which makes this method slower. The mapping admits no fixed point precisely when for some (t, S) the candidate $\rho_{(\rho^0, \rho^1, 0, 1)}$ exhibited in Guyon and Henry-Labordère (2011b) is not PSD.

- $a = v_{\rho^0}$ and $b = v_{\rho^1} - v_{\rho^0}$ (local in index volatility model (Langnau 2010)): in this case, one assumes that the instantaneous variance of the index within model (8.1) is local in index. We denote $\lambda_{(\rho^0, \rho^1, a, b)} = \lambda_{(\rho^0, \rho^1)}^*$:

$$\lambda_{(\rho^0, \rho^1)}^*(t, S_t) \equiv \frac{I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 - v_{\rho^0}(t, S_t)}{v_{\rho^1}(t, S_t) - v_{\rho^0}(t, S_t)}.$$

This is the only situation where no estimation of conditional expectation is needed. Note that $\lambda_{(\rho^0, \rho^1)}^*$ is well defined, even if the corresponding $\rho_{(\rho^0, \rho^1)}^*$ is not PSD, as long as $v_{\rho^1} - v_{\rho^0}$ does not vanish.

Another choice of (a, b) that respects the symmetry of the problem is the following.

- $a \equiv 0$ and $b = v_{\rho^1} - v_{\rho^0}$: in this case,

$$\begin{aligned} & \lambda_{(\rho^0, \rho^1, 0, v_{\rho^1} - v_{\rho^0})}(t, S_t) \\ &= \frac{I_t^2 \sigma_{\text{Dup}}^I(t, I_t)^2 - \mathbb{E}_{\rho_{(\rho^0, \rho^1, 0, v_{\rho^1} - v_{\rho^0})}}[v_{\rho^0}(t, S_t) \mid I_t]}{v_{\rho^1}(t, S_t) - v_{\rho^0}(t, S_t)}. \end{aligned}$$

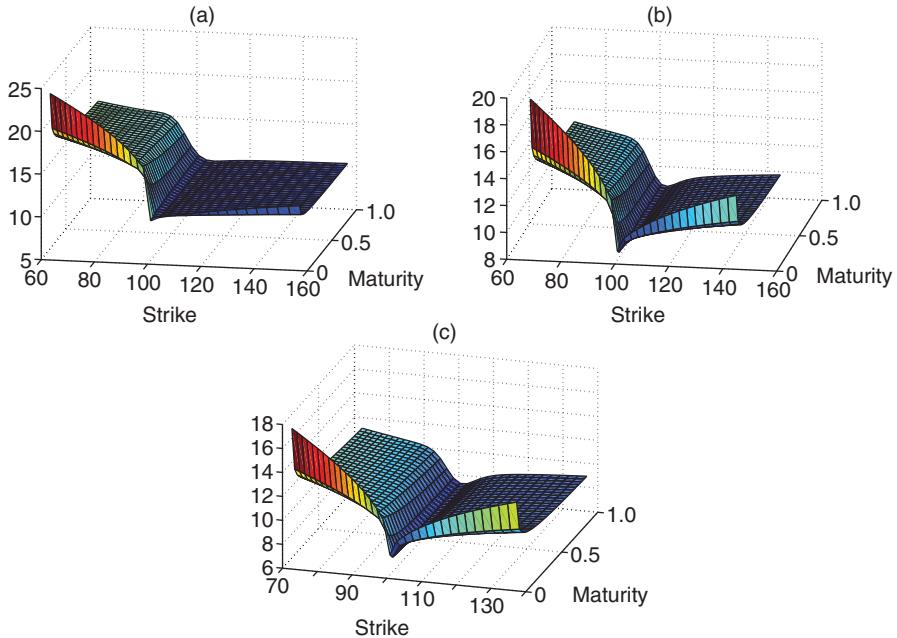
REMARK 8.4 As mentioned in Remark 3.4, our method easily handles local correlations that depend on path-dependent variables. It is enough to add those path-dependent variables to the arguments of the functions ρ^0 , ρ^1 , a , b and λ . It also extends to stochastic volatility, stochastic interest rates and stochastic dividend yields (see Section 10).

9 NUMERICAL EXPERIMENTS

9.1 The foreign exchange triangle smile calibration problem

9.1.1 Calibration

We have tested several local in cross $a + bp$ models on March 2012 market data, involving the three currencies USD, EUR and GBP, and using USD as our domestic currency: $S^1 = \text{EUR}/\text{USD}$, $S^2 = \text{GBP}/\text{USD}$ and $S^{12} = S^1/S^2 = \text{EUR}/\text{GBP}$. For simplicity, we have assumed zero interest rates. The three surfaces of local volatilities are shown in Figure 2. Different pairs of functions (a, b) are tested. The results are shown in Figures 3–15. For the three examples in Section 5 (Figures 3–5), we display

FIGURE 2 Surfaces of local volatilities σ_1 , σ_2 and σ_{12} .(a) Local volatility of S^1 . (b) Local volatility of S^2 . (c) Market local volatility of S^{12} .

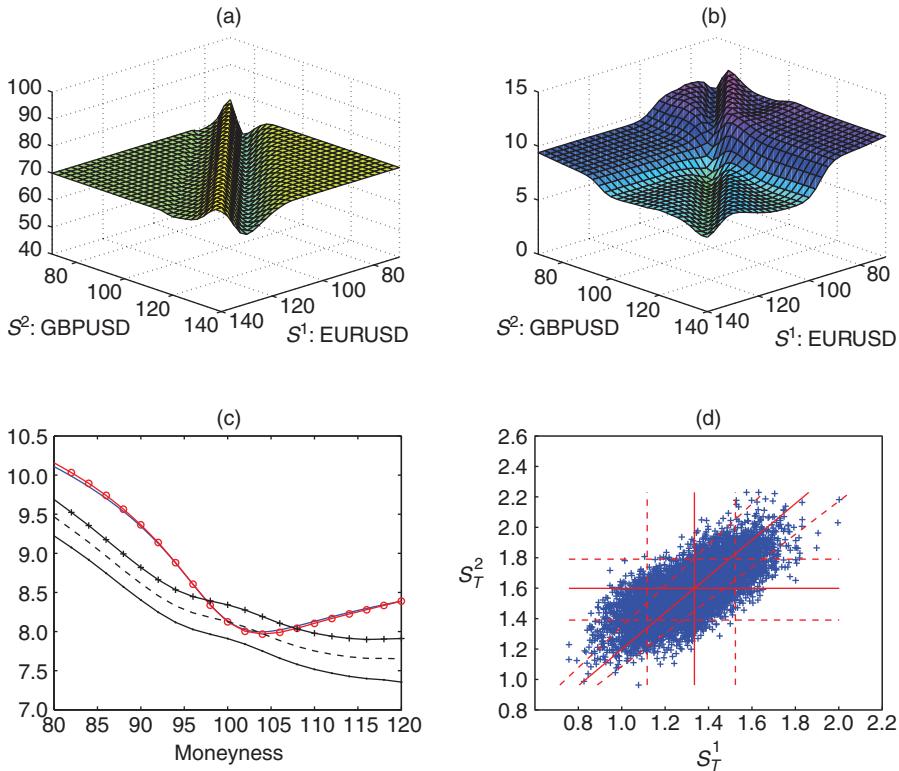
the instantaneous correlation $(S^1, S^2) \mapsto \rho_{(a,b)}(T, S^1, S^2)$ at maturity (part (a)), the instantaneous volatility

$$(S^1, S^2) \mapsto \sqrt{\sigma_1^2(T, S^1) + \sigma_2^2(T, S^2) - 2\rho_{(a,b)}(T, S^1, S^2)\sigma_1(T, S^1)\sigma_2(T, S^2)}$$

of the cross-rate at maturity (part (b)), the repriced smile of the cross-rate at maturity T (part (c)) and the scatter plot of Monte Carlo sampled paths (S_T^1, S_T^2) (part (d)). In the case where the instantaneous correlation $\rho_{(a,b)}(t, S^1, S^2)$ takes values above $+1$, we simply cap it at $+1$. We highlight the corresponding (S_T^1, S_T^2) on the scatter plot with red circles. For other pairs (a, b) , we only display a subset of the four graphs. Full results are available in Guyon (2013).

We picked $T = 1$ and used the particle method described in Section 4 with $N = 10\,000$ Monte Carlo paths and the time step $\Delta t = \frac{1}{80}$. For the nonparametric regressions, we used the quartic kernel $K(x) = (1 - x^2)^2 1_{\{|x| \leq 1\}}$ and a bandwidth

$$h = \kappa \bar{\sigma}^{12} S_0^{12} \sqrt{\max(t, t_{\min})} N^{-1/5},$$

FIGURE 3 $a = 0, b = 1$.

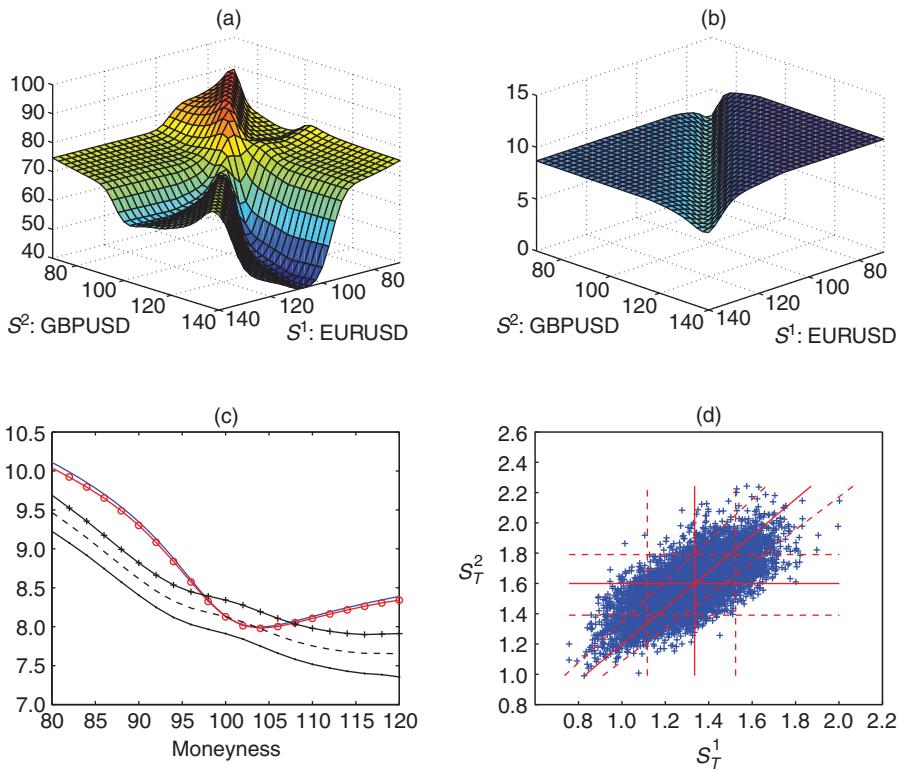
$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . Market (blue line); repriced using calibrated local correlation (red circles); constant correlation: 71% (black crosses); constant correlation: 72% (black dashes); constant correlation: 73% (black circles). (d) 10 000 samples of (S_T^1, S_T^2) (blue crosses); expectation (red line); 10% and 90% quantiles (red dashes).

where $\bar{\sigma}^{12} = 10\%$ is a typical level for the volatility of S^{12} , $t_{\min} = 0.25$ and $\kappa = 3$. The conditional expectations are computed on a grid $G_{S,t}$ of $N_{S,t} = \max(N_S \sqrt{t}, N'_S)$ values of the conditioning random variable $S^{12} = S^1/S^2$, with $N_S = 30$ and $N'_S = 15$. We use the 1% and 99% quantiles of the distribution of S_t^{12} as the minimum and maximum values of the grid $G_{S,t}$. Then, the function

$$f\left(t, \frac{S^1}{S^2}\right) = a(t, S^1, S^2) + b(t, S^1, S^2)\rho(t, S^1, S^2)$$

is interpolated using cubic splines and extrapolated in a flat way.

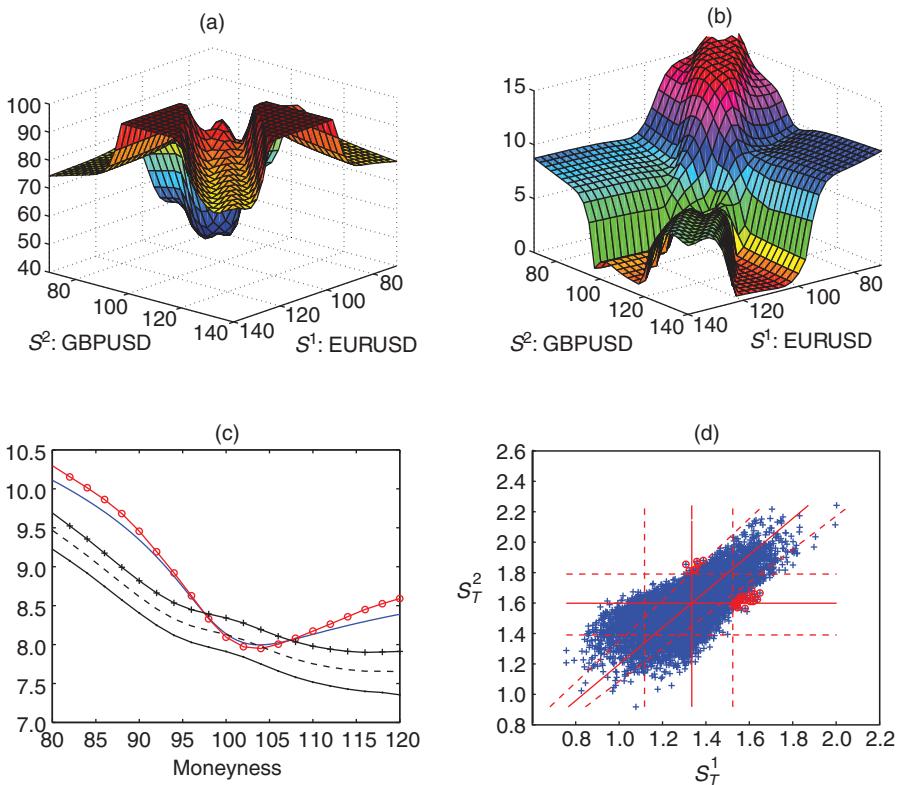
Part (c) in each figure also shows the market implied volatilities of the cross-rate at maturity, as well as the smiles produced by the constant correlation model for three

FIGURE 4 $a(t, S^1, S^2) = \sigma_1^2(t, S^1) + \sigma_2^2(t, S^2)$ and $b(t, S^1, S^2) = -2\sigma_1(t, S^1)\sigma_2(t, S^2)$.

$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . For complete key to parts (c) and (d), see Figure 3.

values of constant correlation: 71%, 72% and 73%. This allows us to translate the calibration error in terms of correlation points. The value of the constant correlation that fits the market value of the at-the-money implied volatility of the cross-rate at maturity is 72%.

Figures 3–13 illustrate the variety of (at least almost) admissible correlations. Before we introduced the local in cross $a + b\rho$ representation, one could only choose between two local correlations: the local in cross correlation $\rho_{(0,1)}$ (Figure 3) or the correlation ρ^* corresponding to a local in cross volatility of the cross (Figure 4). Thanks to our local in cross $a + b\rho$ representation, among the variety of admissible correlations that it produces, one can now pick one's favorite, depending on one's criteria.

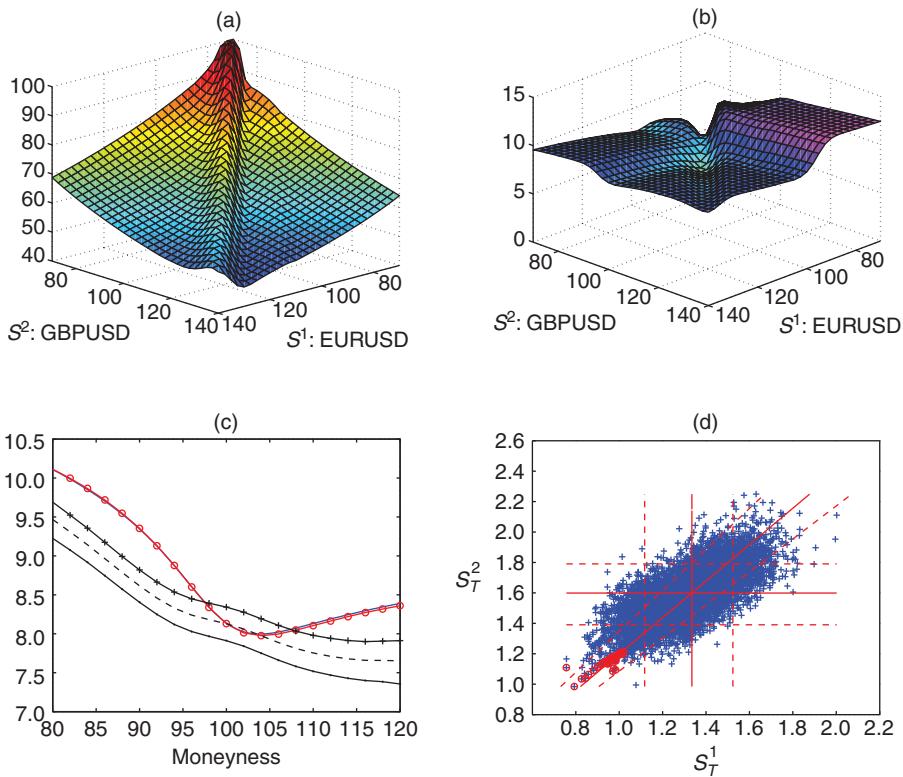
FIGURE 5 $a = 0$ and $b(t, S^1, S^2) = \sigma_1(t, S^1)\sigma_2(t, S^2)$.

$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . Samples where correlation > 1 are shown as red circles. Proportion of sampled paths where correlation was capped at +1: 0.3%. For complete key to parts (c) and (d), see Figure 3.

- Match a view on the correlation skew, for instance, match a value of

$$\Delta\rho \equiv \rho(T, 1.05S_0^1, 1.05S_0^2) - \rho(T, 0.95S_0^1, 0.95S_0^2).$$

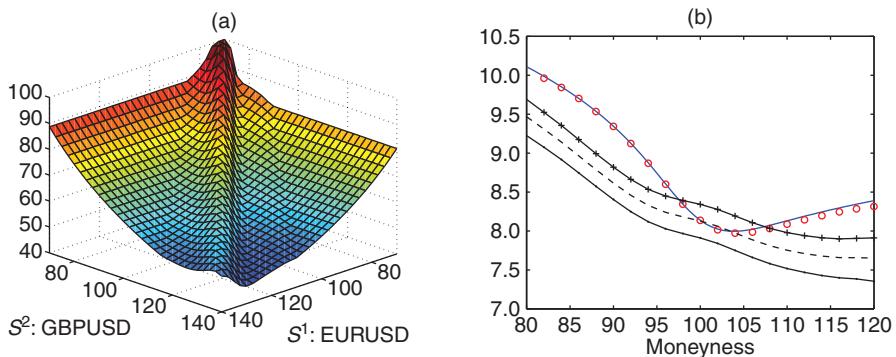
- Reproduce some features of the historical shape of the correlation between the returns in S^1 and S^2 as a function of the values of S^1 and S^2 . Figure 16 shows what this shape looks like for EUR/USD and GBP/USD over the periods January 2007–June 2013 (part (a)) and January 2011–June 2013 (part (b)).
- Fit the price of options on S^1 and S^2 (other than the payoffs $(S^1 - KS^2)_+$, which are automatically fitted to the market, since the correlation is admissible). See Table 2.

FIGURE 6 $a = 0$ and $b(t, S^1, S^2) = \sqrt{S^1 S^2}$.

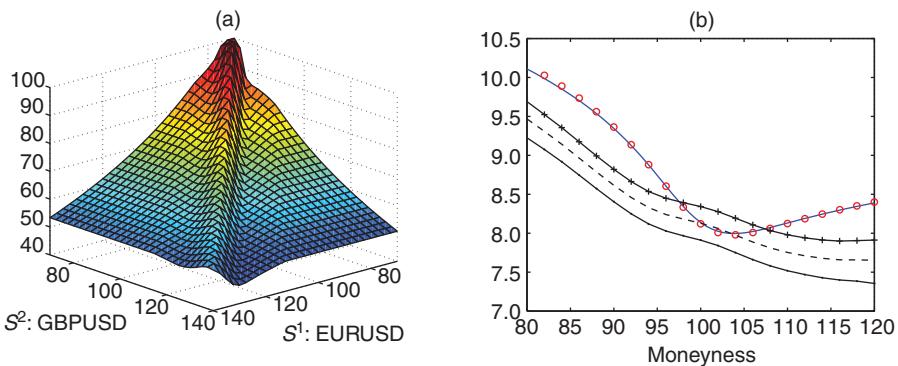
$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . Samples where correlation > 1 are shown as red circles. Proportion of sampled paths where correlation was capped at $+1$: 0.4%. For complete key to parts (c) and (d), see Figure 3.

For instance, one may fit a negative skew by using $a = 0$ and $b = (S^1 S^2)^\alpha$, with $\alpha > 0$ (see Figures 6 and 9). The larger α , the more negative the skew $\Delta\rho$. However, overly large values of α produce correlation candidates that are not admissible. For instance, in Figure 6, we observe that we had to cap the correlation at 1 for some small values of S^1 and S^2 . However, this is still acceptable in practice, because only 0.3% of the simulated spots undergo this capped correlation. Conversely, one may fit a positive skew by using a negative value for α (see Figure 10).

If one observes that the historical correlation is large when the spots are low – which is typical in equity markets – then one may slightly transform function b and use, for instance, $b = \min(S^1, S^2)^{2\alpha}$ (see Figure 7). This has almost no impact on the price of many products (see Table 2), but it allows us to incorporate correlation

FIGURE 7 $a = 0$ and $b(t, S^1, S^2) = \min(S^1/S_0^1, S^2/S_0^2)$.

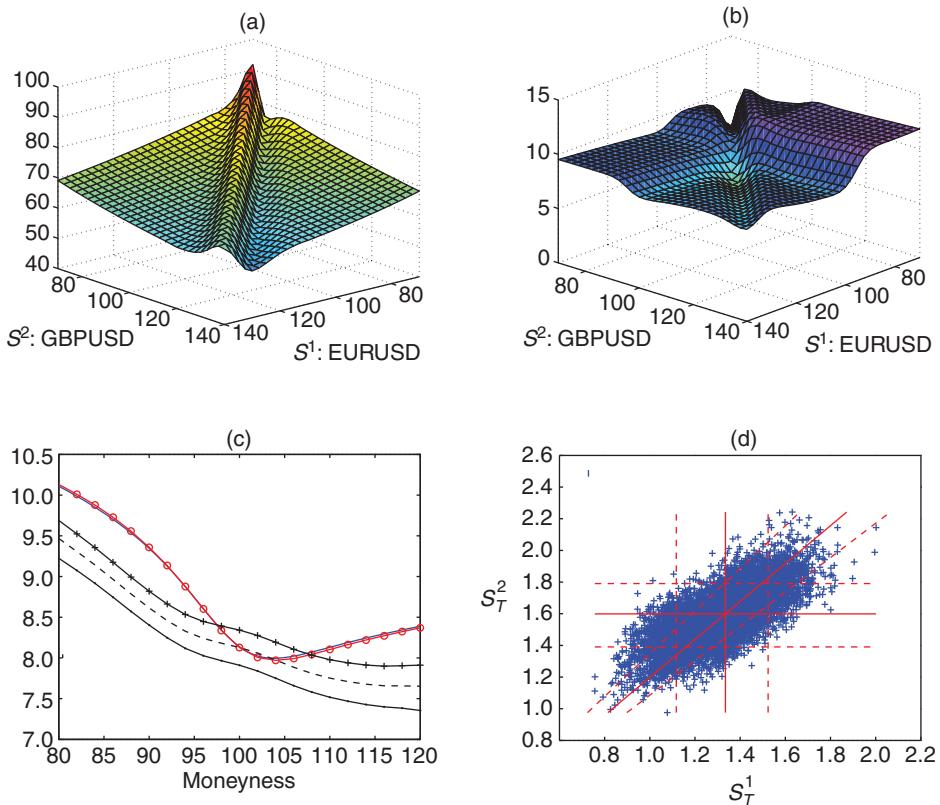
$T = 1$. (a) Local correlation. (b) Implied volatility of S^1/S^2 . Market (blue line); repriced using calibrated local correlation (red circles); constant correlation: 71% (black crosses); constant correlation: 72% (black dashes); constant correlation: 73% (black circles).

FIGURE 8 $a = 0$ and $b(t, S^1, S^2) = \max(S^1/S_0^1, S^2/S_0^2)$.

$T = 1$. (a) Local correlation. (b) Implied volatility of S^1/S^2 . For complete key to part (b), see Figure 7.

impact into the Delta hedge and avoid posting large remarking-to-market loss in the case of a crisis. If, on the contrary, one wants to decrease correlation for low spot values, one may use, for instance, $b = \max(S^1, S^2)^{2\alpha}$ (see Figure 8).

Choosing $\rho_{(0,1)}$ implies pricing vanishing correlation skew across lines where S^1/S^2 is constant and may not be desirable. Choosing ρ^* may imply pricing and hedging with a correlation that varies strongly with the spot values, and which is

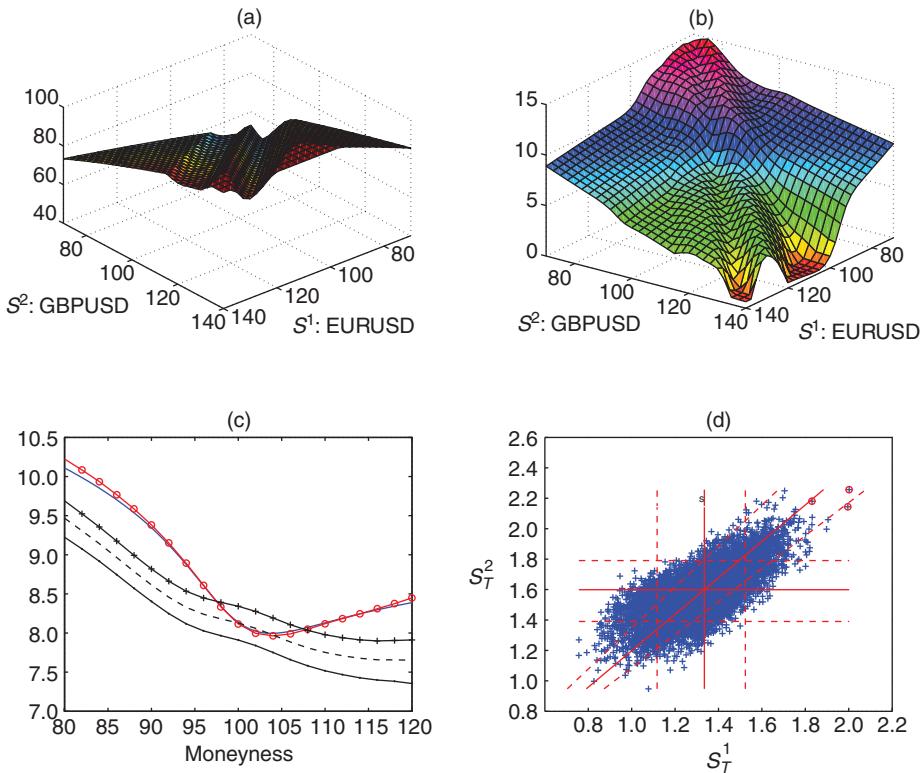
FIGURE 9 $a = 0$ and $b(t, S^1, S^2) = (S^1 S^2)^{1/4}$.

$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . For complete key to parts (c) and (d), see Figure 3.

highly asymmetric (see Figure 4). As expected from (6.1), the image of $\rho_{(0,1)}$ is much narrower than the image of $\rho^*: \rho_{(0,1)}$ varies much less than ρ^* .

Note that, in this numerical example, the third financially natural correlation, namely the local in cross covariance correlation, varies a lot (as does the volatility of S^1/S^2). It has to be capped at 1 for S^1 large and S^2 around the money as well as for S^2 large and S^1 around the money. However, only around 0.3% of the simulated paths are affected by the cap.

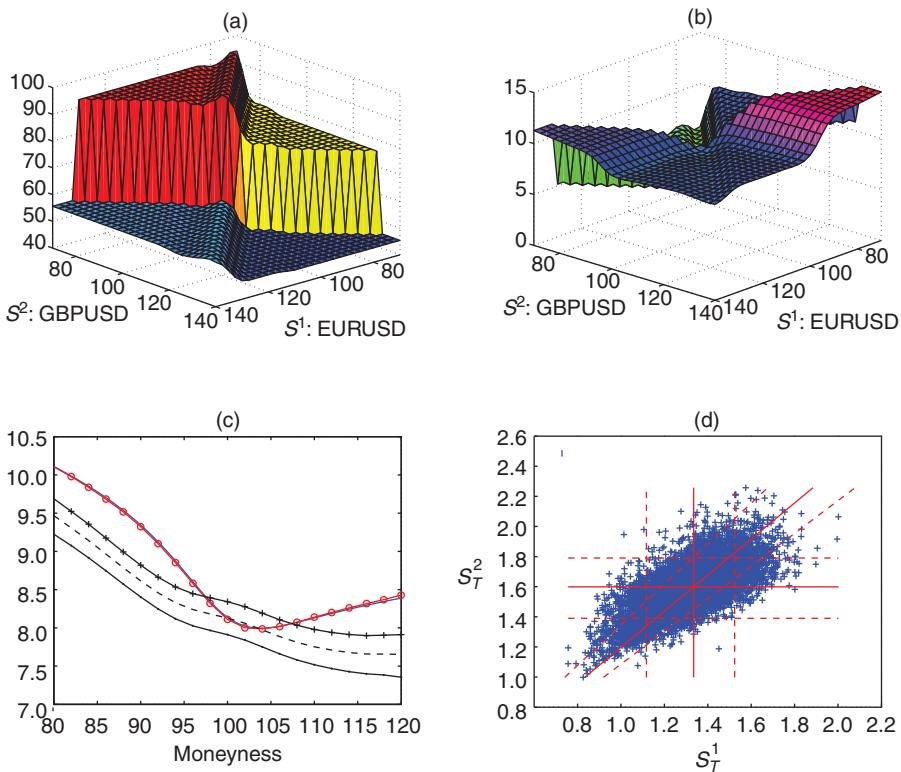
An extreme admissible correlation is shown in Figure 11. From (6.3), we know that all admissible correlations approximately share the same average value on each line where S^1/S^2 is constant. (This is exact only when the two rates have no skew, and

FIGURE 10 $a = 0$ and $b(t, S^1, S^2) = 1/\sqrt{S^1 S^2}$.

$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . Samples where correlation > 1 are shown as red circles. Proportion of sampled paths where correlation was capped at +1: 0.03%. For complete key to parts (c) and (d), see Figure 3.

when the average is taken under \mathbb{Q}^f .) The correlation in Figure 11 was built so that the local correlation is very high and roughly constant when $(S^1/S_0^1) + (S^2/S_0^2)$ is less than 2, and very low and roughly constant when $(S^1/S_0^1) + (S^2/S_0^2)$ is greater than 2; it also has the correct average value on those lines where the cross is constant. A smoothed version using the tanh function is shown in Figure 12.

Finally, Figure 13 shows that our new method allows us to build very diverse admissible correlations: in this instance, a local correlation that is peaked around (S_0^1, S_0^2) . As for Figures 14 and 15, they illustrate that a wrong choice of functions (a, b) can lead to inadmissible correlations, with a high proportion (respectively 22% and 16%) of correlations that have to be capped. This results in a poor calibration of the smile of the cross-rate.

FIGURE 11 $a = 0$ and $b(t, S^1, S^2) = 1.5 + \mathbf{1}_{\{(S^1/S_0^1)+(S^2/S_0^2)>2\}} \cdot$ 

$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . For complete key to parts (c) and (d), see Figure 3.

9.1.2 Pricing

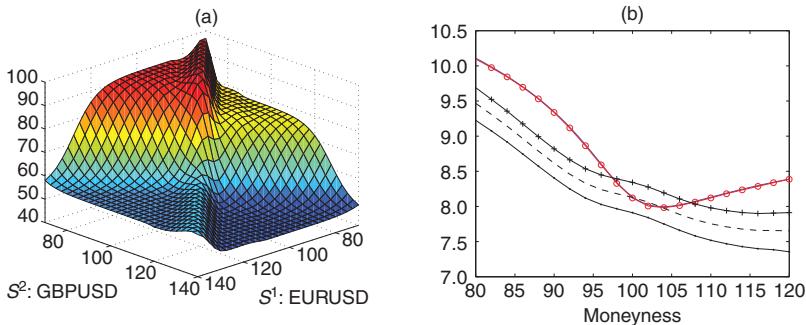
To illustrate the impact of the local correlation model on the price of options, we have considered the following three derivative products.

Min of calls:

$$g(S_T^1, S_T^2) = \min \left(\left(\frac{S_T^1}{K^1} - 1 \right)_+, \left(\frac{S_T^2}{K^2} - 1 \right)_+ \right), \quad K^1 = S_0^1, \quad K^2 = S_0^2.$$

Put on worst:

$$g(S_T^1, S_T^2) = \left(K - \min \left(\frac{S_T^1}{S_0^1}, \frac{S_T^2}{S_0^2} \right) \right)_+, \quad K = 0.95.$$

FIGURE 12 $a = 0$ and $b(t, S^1, S^2) = 1.5 + \frac{1}{2}(1 + \tanh(10((S^1/S_0^1) + (S^2/S_0^2) - 2)))$.

$T = 1$. (a) Local correlation. (b) Implied volatility of S^1/S^2 . For complete key to part (b), see Figure 7.

Put on basket:

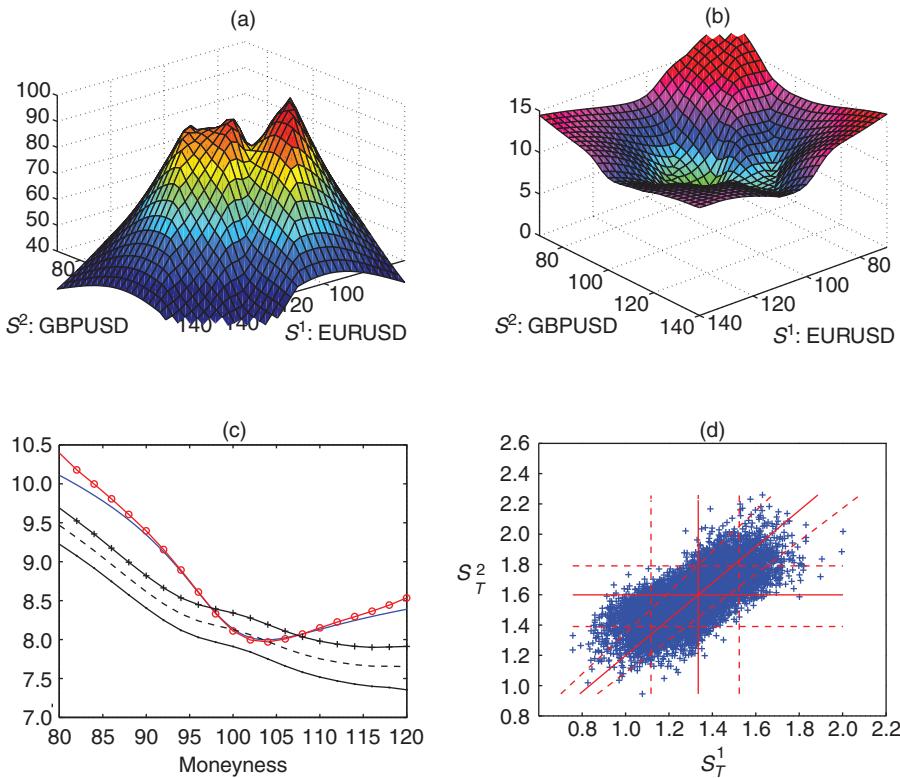
$$g(S_T^1, S_T^2) = \left(K - \left(\frac{S_T^1}{S_0^1} + \frac{S_T^2}{S_0^2} \right) \right)_+, \quad K = 1.8.$$

The prices are shown in Table 2 for different admissible correlations. For each of these products, we can build an intuition of the impact of the local correlation model on the price by looking at the instantaneous correlation surfaces in Figures 3–12 and using the price impact formula (7.1). To this end, note that the cross Gammas of these options at maturity are simply proportional to the following Dirac masses:

$$\begin{aligned} \text{Min of calls: } & \delta\left(\frac{S^2}{K^2} - \frac{S^1}{K^1}\right) \mathbf{1}_{\{S^1/K^1 \geq 1\}}, \\ \text{Put on worst: } & -\delta\left(\frac{S^2}{S_0^2} - \frac{S^1}{S_0^1}\right) \mathbf{1}_{\{S_T^1/S_0^1 \leq K\}}, \\ \text{Put on basket: } & \delta\left(\frac{S^1}{S_0^1} + \frac{S^2}{S_0^2} - K\right). \end{aligned}$$

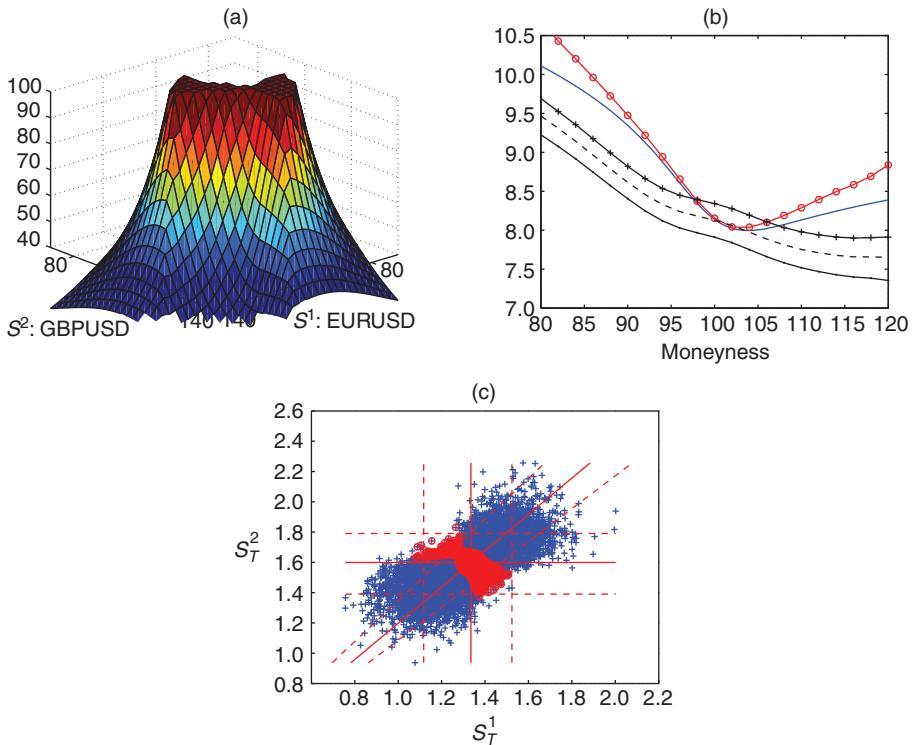
From (7.1), we expect the higher prices to correspond to local correlation surfaces that are

- larger in the neighborhood of the half-line $S^1/S^2 = K^1/K^2$, $S^1 \geq K^1$, for the min of calls,
- smaller in the neighborhood of the half-line $S^1/S^2 = S_0^1/S_0^2$, $S^1 \leq K S_0^1$, for the put on worst,
- larger in the neighborhood of the segment $(S^1/S_0^1) + (S^2/S_0^2) = K$, $S^1, S^2 > 0$, for the put on basket.

FIGURE 13 $a = 0$ and $b(t, S^1, S^2) = 0.08 + ((S^1/S_0^1) - 1)^2 + ((S^2/S_0^2) - 1)^2$.

$T = 1$. (a) Local correlation. (b) Local volatility of S^1/S^2 . (c) Implied volatility of S^1/S^2 . (d) 10 000 samples of (S_T^1, S_T^2) . For complete key to parts (c) and (d), see Figure 3.

This is indeed verified. For these products, the highest and lowest prices always correspond to the local in cross covariance correlation (Figure 5) as well as to the extreme correlation of Figure 11. This way, we have provided a numerical partial answer to the difficult problem of determining the lower and upper bounds of prices of options on (S^1, S^2) given the three surfaces of implied volatilities on S^1 , S^2 and S^1/S^2 , and the corresponding models. The answer is only partial because we have only considered local volatility models (see Section 10 for a generalization to stochastic local volatility models and Guyon (2016) for the generalization to cross-dependent volatility models) and a small number of admissible local correlations $\rho_{(a,b)}$. Note that the range of prices is already quite large, from 2.37 to 2.91, for instance, for the min of calls, despite the fact that the three surfaces of implied volatilities are calibrated. In Table 2,

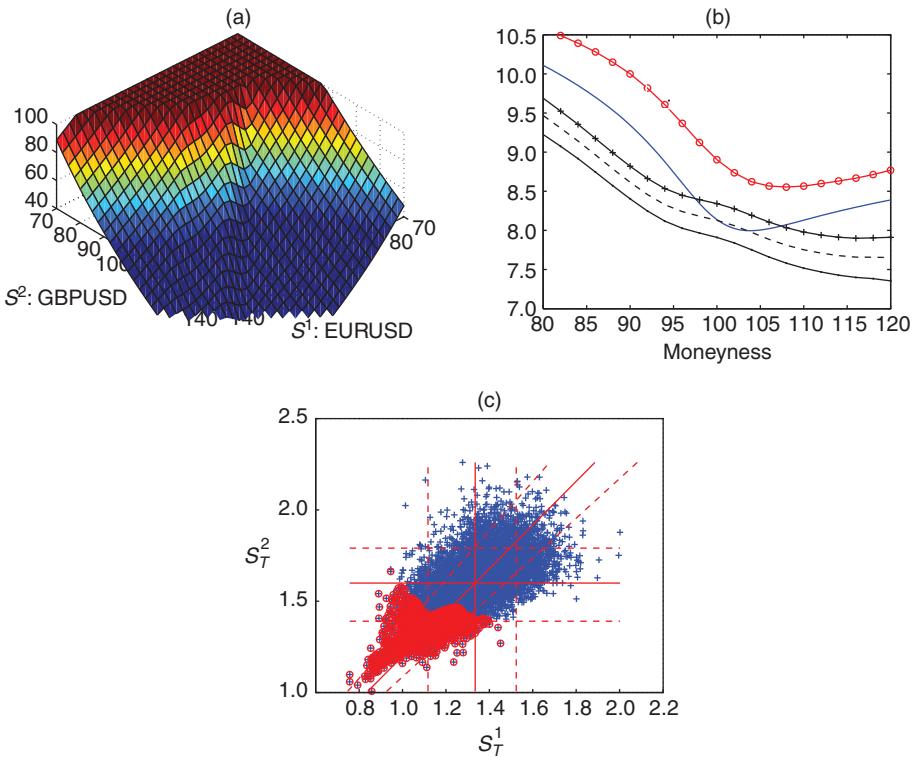
FIGURE 14 $a = 0$ and $b(t, S^1, S^2) = 0.02 + ((S^1/S_0^1) - 1)^2 + ((S^2/S_0^2) - 1)^2$.

$T = 1$. (a) Local correlation. (b) Implied volatility of S^1/S^2 . (c) 10 000 samples of (S_T^1, S_T^2) . Samples where correlation > 1 are shown as red circles. Proportion of sampled paths where correlation was capped at $+1$: 22%. For complete key to parts (b) and (c), see Figure 3.

we have also reported prices of a digital call on the cross-rate with (American) barrier $K = 1.1(S_0^1/S_0^2)$ and a double-no-touch on the cross-rate with (American) barriers $K_1 = 0.9(S_0^1/S_0^2)$ and $K_2 = 1.1(S_0^1/S_0^2)$. The derivation of lower and upper bounds of prices of calls on the cross-rate S^1/S^2 of maturity T , given the two smiles at maturity T of S^1 and S^2 and the at-the-money implied volatility and skew of S^1/S^2 , can be found in Henry-Labordère (2011).

9.2 The equity index smile calibration problem

We now turn to the equity index smile calibration problem described in Section 8. Here, we test several local in cross $a + b\rho$ models on synthetic data involving $N = 10$ stocks. For simplicity, we used $\rho^0 = \text{Id}$ and $\rho^1 = \mathbf{1}$, and we assumed zero interest

FIGURE 15 $a(t, S^1, S^2) = 3\sqrt{S^1 S^2}$ and $b(t, S^1, S^2) = \sqrt{S^1 S^2}$.

$T = 1$. (a) Local correlation. (b) Implied volatility of S^1/S^2 . (c) 10 000 samples of (S_T^1, S_T^2) . Samples where correlation > 1 are shown as red circles. Proportion of sampled paths where correlation was capped at $+1$: 16%. For complete key to parts (b) and (c), see Figure 3.

rates, repo and dividends; flat surfaces of implied volatility at 20% and equal initial value $S_0^i = 100$ for the N stocks; uniform index weights $\alpha_i = 1/N$; and a simple time-homogeneous local volatility for the index:

$$\sigma_{\text{Dup}}^I(t, I) = \bar{\sigma} - (\bar{\sigma} - \underline{\sigma}) \left(\frac{1}{2} + \frac{1}{2} \tanh \left(s \left(\frac{I}{I_0} - 1 \right) \right) \right), \quad (9.1)$$

with $\underline{\sigma} = 9\%$, $\bar{\sigma} = 19\%$ and the at-the-money skew parameter $s = 40$; σ_{Dup}^I decreases from $\bar{\sigma}$ to $\underline{\sigma}$ as I increases, and the larger s , the steeper the skew (see Figure 17(a)). Table 3 shows the prices of puts and calls on worst, with respective payoffs $(K - \min_i(S_T^i/S_0^i))_+$ and $(\min_i(S_T^i/S_0^i) - K)_+$, $K = 70\%$, and puts on best $(K - \max_i(S_T^i/S_0^i))_+$, $K = 100\%$, in various local correlation models indexed

FIGURE 16 Empirical state-dependency of correlation of EUR/USD and GBP/USD over the periods (a) January 2007–June 2013 and (b) January 2011–June 2013.

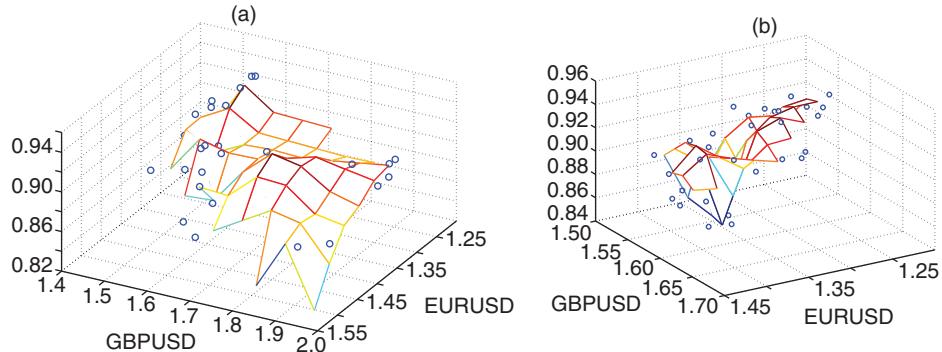
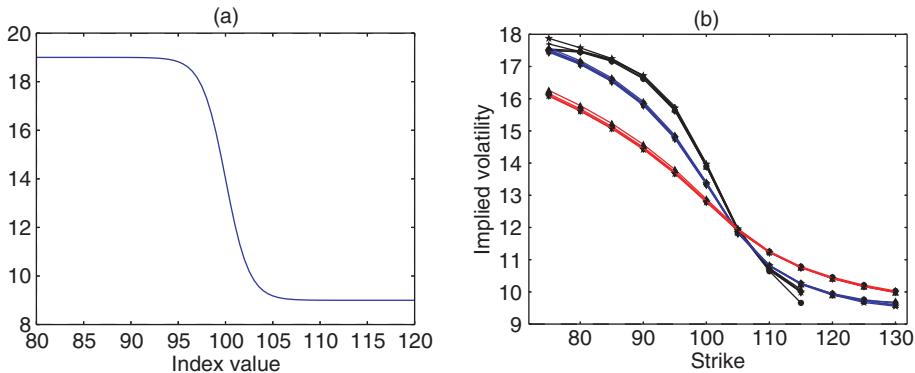


TABLE 2 Price in pct of the min of calls, put on worst, put on basket, digital call (DC) on S^{12} , and double-no-touch (DNT) on S^{12} for different admissible (or almost admissible) correlations described by the pair of functions (a, b) .

<i>a</i>	<i>b</i>	Min of calls	Put on worst	Put on basket	DC on S^{12}	DNT on S^{12}
Standard deviation		≈0.020	≈0.027	≈0.027	≈0.18	≈0.22
Constant correlation 72%		2.59	3.47	1.88	21.18	57.97
0	1	2.65	3.49	1.91	20.53	58.02
$\sigma_1^2 + \sigma_2^2$	$-2\sigma_1\sigma_2$	2.53	3.37	1.99	20.46	58.17
0	$\sigma_1\sigma_2$	2.91	3.70	1.78	19.75	59.41
0	$\sqrt{S^1 S^2}$	2.56	3.41	1.95	20.48	58.11
0	$\max(S^1, S^2)$	2.56	3.40	1.95	20.50	58.10
0	$\min(S^1, S^2)$	2.56	3.41	1.95	20.46	58.15
0	$(S^1 S^2)^{1/4}$	2.61	3.45	1.93	20.45	58.09
0	$\sqrt{\max(S^1, S^2)}$	2.61	3.45	1.93	20.46	58.08
0	$\sqrt{\min(S^1, S^2)}$	2.61	3.45	1.93	20.44	58.10
0	$\frac{1}{\sqrt{S^1 S^2}}$	2.74	3.56	1.87	20.41	58.21
0	$1.5 + 1_{\{(S^1/S_0^1)+(S^2/S_0^2)>2\}}$	2.37	3.25	2.06	19.71	59.26
0	$2 + \frac{1}{2} \tanh(10((S^1/S_0^1) + (S^2/S_0^2) - 2))$	2.42	3.28	2.04	20.14	58.65

We used the same $50\,000 = 10\,000 + 40\,000$ Brownian paths for all choices of (a, b) . The first 10 000 paths are those used for calibration of the correlation.

FIGURE 17 Local volatility of index, and market and calibrated smiles of index.

(a) Time-homogeneous local volatility of the index. (b) Corresponding “market” smiles and smiles of calibrated models for $\rho^0 = \text{Id}$, $\rho^1 = \mathbf{1}$, and various pairs (a, b) , with $a = 0$, corresponding to the extremal prices of put on worst, call on worst and put on best in Table 3. $T = 1/4$ for all black lines: market (black triangles); calibration, $b = \min^{-3/2}$ (black circles); calibration, $b = \min^{-2}$ (black stars); calibration, $b = \max^{-2}$ (black diamonds); calibration, $b = \max^{-2}$ (black crosses), where \min and \max respectively denote $\min_i(S^i/S_0^i)$ and $\max_i(S^i/S_0^i)$. $T = 1$ for all blue lines. Same key as for black lines. $T = 4$ for all red lines. Same key as for black lines. We used 4000 paths and twenty time steps per year. Computation time: 0.37 seconds ($T = 0.25$), 1.5 seconds ($T = 1$) and 6.3 seconds ($T = 4$), using Python on a single processor Intel Core i5-3570 CPU @ 3.40GHz with 8 GB of RAM.

by (a, b) , with $a = 0$ for maturity $T = 4$ years. All the models in Table 3 are admissible, ie, are calibrated to the full surfaces of implied volatility of the N stocks and the index up to $T = 4$. Within this family of models, the lower and upper bounds of options on worst correspond to $b = (\min(S^i/S_0^i))^p$ for $p \in \{-3/2, 2\}$; those of the put on best correspond to $b = (\max(S^i/S_0^i))^p$ for $p \in \{-2, 2\}$. Those bounds help us quantify model risk within the family of local correlation models. Even within this restricted class, model risk is already high: the range of prices is large for the put on worst ($11.44 - 12.46$, a 9% difference), and even larger for the call on worst ($1.89 - 2.80$, 48%) and the put on best ($3.87 - 5.37$, 39%). In Figure 17(b), one can verify that the smiles for these four models are indeed calibrated to the “market” smile corresponding to (9.1).

10 GENERALIZATION TO STOCHASTIC VOLATILITY, STOCHASTIC INTEREST RATES AND STOCHASTIC DIVIDEND YIELD

10.1 The foreign exchange triangle smile calibration problem

We now generalize the construction of families of local correlation models for the FX triangle smile calibration problem in the presence of local stochastic volatility,

TABLE 3 Price in pct of the put on worst, call on worst and put on best for different admissible correlations described by the pair of functions (a, b) , when $T = 4$.

a	b	Put on worst, $K = 70$	Call on worst, $K = 70$	Put on best, $K = 100$
Standard deviation		≈0.06	≈0.03	≈0.06
Constant correlation 34%		14.86	3.20	1.14
0	1	11.90	2.31	4.62
0	$(\prod_i (S^i / S_0^i))^{-1}$	12.43	2.39	4.10
0	$(\prod_i (S^i / S_0^i))^{-1/2}$	12.11	2.37	4.42
0	$(\prod_i (S^i / S_0^i))^{1/2}$	11.87	2.22	4.60
0	$\prod_i (S^i / S_0^i)$	12.06	2.10	4.35
0	$1 + 1_{\{S^1 / S_0^1 \leq S^2 / S_0^2\}}$	11.81	2.54	4.98
0	$1 + 2 \times 1_{\{S^1 / S_0^1 \leq S^2 / S_0^2\}}$	11.95	2.55	5.01
0	$(\min(S^i / S_0^i))^{-3/2}$	11.44	2.80	5.02
0	$(\min(S^i / S_0^i))^{-1}$	11.65	2.63	4.89
0	$\min(S^i / S_0^i)$	12.17	2.07	4.45
0	$(\min(S^i / S_0^i))^{3/2}$	12.30	1.98	4.33
0	$(\min(S^i / S_0^i))^2$	12.46	1.89	4.19
0	$\max(S^i / S_0^i)^{-2}$	12.18	2.08	3.87
0	$(\max(S^i / S_0^i))^{-3/2}$	12.09	2.14	4.03
0	$(\max(S^i / S_0^i))^{-1}$	12.01	2.20	4.20
0	$\max(S^i / S_0^i)$	11.76	2.48	5.19
0	$(\max(S^i / S_0^i))^2$	11.87	2.62	5.37

stochastic interest rates and local correlation. For simplicity (see Remark 10.5 for a more general case), let us assume that the extra Brownian motions W^3, W^4, W^5, \dots that drive the dynamics of the stochastic volatilities and the stochastic interest rates are independent of the two Brownian motions W^1 and W^2 that drive the dynamics of (S^1, S^2) . The correlation matrix of W^3, W^4, W^5, \dots is assumed to be known and constant. Only the correlation between W^1 and W^2 is unknown; it is assumed to be local:

$$\begin{aligned} dS_t^1 &= (r_t^d - r_t^1)S_t^1 dt + \sigma_1(t, S_t^1)a_t^1 S_t^1 dW_t^1, \\ dS_t^2 &= (r_t^d - r_t^2)S_t^2 dt + \sigma_2(t, S_t^2)a_t^2 S_t^2 dW_t^2, \\ d\langle W^1, W^2 \rangle_t &= \rho(t, S_t^1, S_t^2, a_t^1, a_t^2, D_{0t}^d, D_{0t}^1, D_{0t}^2) dt. \end{aligned} \quad (10.1)$$

$a_t^1, a_t^2, r_t^d, r_t^1$ and r_t^2 are Ito processes driven by the extra Brownian motions W^3, W^4, W^5, \dots . The instantaneous correlation between W^1 and W^2 is assumed to depend not only on the FX rates S_t^1 and S_t^2 but also on the stochastic volatilities a_t^1 and a_t^2 as well

as on the (stochastic) discount factors D_{0t}^d , D_{0t}^1 and D_{0t}^2 : $D_{0t}^i = \exp(-\int_0^t r_s^i ds)$.² To keep our notation short, we write $\rho(t, X_t)$ with $X_t = (S_t^1, S_t^2, a_t^1, a_t^2, D_{0t}^d, D_{0t}^1, D_{0t}^2)$.

First, the local volatility $\sigma_1(t, S^1)$ is calibrated to the market smile of S^1 , using Proposition 22 in the online appendix, with $r_t = r_t^d$, $q_t = r_t^1$ and $a_t = \sigma_1(t, S_t^1)a_t^1$:

$$\begin{aligned} \sigma_1(t, K)^2 & \frac{\mathbb{E}^{\mathbb{Q}}[D_{0t}^d(a_t^1)^2 | S_t^1 = K]}{\mathbb{E}^{\mathbb{Q}}[D_{0t}^d | S_t^1 = K]} \\ &= \sigma_{\text{Dup}}^1(t, K)^2 - \frac{\mathbb{E}^{\mathbb{Q}}[D_{0t}^d(r_t^d - r_t^1 - (r_t^{d,0} - r_t^{1,0}))1_{S_t^1 > K}]}{\frac{1}{2}K\partial_K^2\mathcal{C}_1(t, K)} \\ &\quad + \frac{\mathbb{E}^{\mathbb{Q}}[D_{0t}^d(r_t^1 - r_t^{1,0})(S_t^1 - K)^+]}{\frac{1}{2}K^2\partial_K^2\mathcal{C}_1(t, K)}, \end{aligned}$$

where $r_t^{d,0} = -\partial_t \ln P_{0t}^d$, $r_t^{1,0} = -\partial_t \ln P_{0t}^1$ and

$$\sigma_{\text{Dup}}^1(t, K)^2 = \frac{\partial_t \mathcal{C}_1(t, K) + (r_t^{d,0} - r_t^{1,0})K\partial_K \mathcal{C}_1(t, K) + r_t^{1,0}\mathcal{C}_1(t, K)}{\frac{1}{2}K^2\partial_K^2\mathcal{C}_1(t, K)},$$

with $\mathcal{C}_1(t, K)$ the market price of the call option on S^1 , with strike K and maturity t ; the same holds for $\sigma_2(t, S^2)$. This is achieved using the particle algorithm (see Guyon and Henry-Labordère 2011b; 2014). Knowledge of the local correlation $\rho(t, X)$ between the two FX rates is not required at this step. Then, $\rho(t, X)$ is calibrated to the market smile of the cross-rate S^{12} by using the following proposition (see proof in the online appendix).

PROPOSITION 10.1 *The local volatilities σ_1 and σ_2 being fixed, model (10.1) is calibrated to the market smile of the cross-FX rate if and only if*

$$\begin{aligned} & \frac{\mathbb{E}_{\rho}^{\mathbb{Q}^f}[D_{0t}^2((\sigma_1(t, S_t^1)a_t^1)^2 + (\sigma_2(t, S_t^2)a_t^2)^2 \\ & \quad - 2\rho(t, X_t)\sigma_1(t, S_t^1)a_t^1\sigma_2(t, S_t^2)a_t^2) | S_t^{12} = K]}{\mathbb{E}_{\rho}^{\mathbb{Q}^f}[D_{0t}^2 | S_t^{12} = K]} \\ &= \sigma_{\text{Dup}}^{12}(t, K)^2 - \frac{\mathbb{E}_{\rho}^{\mathbb{Q}^f}[D_{0t}^2(r_t^2 - r_t^1 - (r_t^{2,0} - r_t^{1,0}))1_{S_t^{12} > K}]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t, K)} \\ &\quad + \frac{\mathbb{E}_{\rho}^{\mathbb{Q}^f}[D_{0t}^2(r_t^1 - r_t^{1,0})(S_t^{12} - K)^+]}{\frac{1}{2}K^2\partial_K^2\mathcal{C}(t, K)} \end{aligned} \tag{10.2}$$

² The instantaneous correlation may also depend on r_t^d , r_t^1 , r_t^2 . It may actually depend on any \mathcal{F}_t -measurable random variable, including path-dependent variables (see Remark 3.4).

for all (t, K) , where $D_{0t}^2 = \exp(-\int_0^t r_s^2 ds)$, $r_t^{1,0}$ and $r_t^{2,0}$ are deterministic interest rates, and

$$\sigma_{\text{Dup}}^{12}(t, K)^2 = \frac{\partial_t \mathcal{C}(t, K) + (r_t^{2,0} - r_t^{1,0})K\partial_K \mathcal{C}(t, K) + r_t^{1,0}\mathcal{C}(t, K)}{\frac{1}{2}K^2\partial_K^2 \mathcal{C}(t, K)}$$

is the market local volatility of the cross-rate S^{12} computed using the deterministic interest rates $r_t^{1,0}$ and $r_t^{2,0}$. $\mathcal{C}(t, K)$ is the market price of the call option on S^{12} , with strike K and maturity t .

Equation (10.2) is equivalent to

$$\begin{aligned} & \frac{\mathbb{E}_{\rho}^{\mathbb{Q}}[D_{0t}^d S_t^2 ((\sigma_1(t, S_t^1)a_t^1)^2 + (\sigma_2(t, S_t^2)a_t^2)^2 \\ & \quad - 2\rho(t, X_t)\sigma_1(t, S_t^1)a_t^1\sigma_2(t, S_t^2)a_t^2) | S_t^{12} = K]}{\mathbb{E}_{\rho}^{\mathbb{Q}}[D_{0t}^d S_t^2 | S_t^{12} = K]} \\ &= \sigma_{\text{Dup}}^{12}(t, K)^2 - \frac{\mathbb{E}_{\rho}^{\mathbb{Q}}[D_{0t}^d (S_t^2/S_0^2)(r_t^2 - r_t^1 - (r_t^{2,0} - r_t^{1,0}))1_{S_t^{12} > K}]}{\frac{1}{2}K\partial_K^2 \mathcal{C}(t, K)} \\ & \quad + \frac{\mathbb{E}_{\rho}^{\mathbb{Q}}[D_{0t}^d (S_t^2/S_0^2)(r_t^1 - r_t^{1,0})(S_t^{12} - K)^+]}{\frac{1}{2}K^2\partial_K^2 \mathcal{C}(t, K)} \end{aligned} \quad (10.3)$$

and to

$$\begin{aligned} & \mathbb{E}_{\rho}^{\mathbb{Q}^{f,t}}[(\sigma_1(t, S_t^1)a_t^1)^2 + (\sigma_2(t, S_t^2)a_t^2)^2 \\ & \quad - 2\rho(t, X_t)\sigma_1(t, S_t^1)a_t^1\sigma_2(t, S_t^2)a_t^2 | S_t^{12} = K] \\ &= \sigma_{\text{Dup}}^{12}(t, K)^2 - P_{0t}^2 \frac{\mathbb{E}_{\rho}^{\mathbb{Q}^{f,t}}[(r_t^2 - r_t^1 - (r_t^{2,0} - r_t^{1,0}))1_{S_t^{12} > K}]}{\frac{1}{2}K\partial_K^2 \mathcal{C}(t, K)} \\ & \quad + P_{0t}^2 \frac{\mathbb{E}_{\rho}^{\mathbb{Q}^{f,t}}[(r_t^1 - r_t^{1,0})(S_t^{12} - K)^+]}{\frac{1}{2}K^2\partial_K^2 \mathcal{C}(t, K)}, \end{aligned} \quad (10.4)$$

where $\mathbb{Q}^{f,t}$ denotes the foreign t -forward measure, $d\mathbb{Q}^{f,t}/d\mathbb{Q}^f = D_{0t}^2/P_{0t}^2$.

DEFINITION 10.2 We say that a correlation ρ is admissible if (10.2), or equivalently (10.3) or (10.4), holds.

To build the set of all admissible correlations, we can easily extend Proposition 3.2.

PROPOSITION 10.3 (Local in cross affine transform representation of admissible correlations) Let $\rho \in \mathcal{C}$. It is admissible if and only if there exist two functions a and b of t and $X = (S^1, S^2, a^1, a^2, D_0^d, D_0^1, D_0^2)$ such that b does not vanish and $\rho(t, X) = (f(t, S^{12}) - a(t, X))/b(t, X)$, with $f(t, S^{12}) \equiv N_f(t, S^{12})/D_f(t, S^{12})$

defined by

$$\begin{aligned}
N_f(t, K) &= \frac{1}{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d S_t^2 \mid S_t^{12} = K]} \\
&\quad \times \left(\mathbb{E}_\rho^\mathbb{Q} \left[D_{0t}^d S_t^2 ((\sigma_1(t, S_t^1) a_t^1)^2 + (\sigma_2(t, S_t^2) a_t^2)^2 \right. \right. \\
&\quad \left. \left. + 2 \frac{a(t, X_t)}{b(t, X_t)} \sigma_1(t, S_t^1) a_t^1 \sigma_2(t, S_t^2) a_t^2 \mid S_t^{12} = K \right] \right) \\
&\quad - \sigma_{\text{Dup}}^{12}(t, K)^2 + \frac{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d (S_t^2 / S_0^2) (r_t^2 - r_t^1 - (r_t^{2,0} - r_t^{1,0})) 1_{S_t^{12} > K}]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t, K)} \\
&\quad - \frac{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d (S_t^2 / S_0^2) (r_t^1 - r_t^{1,0}) (S_t^{12} - K)^+]}{\frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t, K)}, \\
D_f(t, K) &= 2 \frac{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d S_t^2 (\sigma_1(t, S_t^1) a_t^1 \sigma_2(t, S_t^2) a_t^2 / b(t, X_t)) \mid S_t^{12} = K]}{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d S_t^2 \mid S_t^{12} = K]}. \quad (10.5)
\end{aligned}$$

PROOF We can always pick two functions $a(t, X)$ and $b(t, X)$ such that b does not vanish and $a(t, X) + b(t, X)\rho(t, X)$ is local in cross. For instance, choose $b \equiv 1$ and $a(t, X) = f(t, S^{12}) - \rho(t, X)$ for some function f . Then, (10.3) is equivalent to

$$\begin{aligned}
&\frac{1}{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d S_t^2 \mid S_t^{12} = K]} \\
&\quad \times \left(\mathbb{E}_\rho^\mathbb{Q} \left[D_{0t}^d S_t^2 \left((\sigma_1(t, S_t^1) a_t^1)^2 + (\sigma_2(t, S_t^2) a_t^2)^2 \right. \right. \right. \\
&\quad \left. \left. \left. + 2 \frac{a(t, X_t)}{b(t, X_t)} \sigma_1(t, S_t^1) a_t^1 \sigma_2(t, S_t^2) a_t^2 \right) \mid S_t^{12} = K \right] \right) \\
&\quad - 2(a + b\rho)(t, K) \frac{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d S_t^2 (\sigma_1(t, S_t^1) a_t^1 \sigma_2(t, S_t^2) a_t^2 / b(t, X_t)) \mid S_t^{12} = K]}{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d S_t^2 \mid S_t^{12} = K]} \\
&= \sigma_{\text{Dup}}^{12}(t, K)^2 - \frac{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d (S_t^2 / S_0^2) (r_t^2 - r_t^1 - (r_t^{2,0} - r_t^{1,0})) 1_{S_t^{12} > K}]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t, K)} \\
&\quad + \frac{\mathbb{E}_\rho^\mathbb{Q}[D_{0t}^d (S_t^2 / S_0^2) (r_t^1 - r_t^{1,0}) (S_t^{12} - K)^+]}{\frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t, K)},
\end{aligned}$$

from which one gets (10.5). Conversely, if two functions (a, b) satisfy (10.5), then ρ satisfies (10.3) and is thus admissible. \square

We denote by $\rho_{(a,b)}$ a solution of (10.5). One can compute $\rho_{(a,b)}$ using the particle method. One can try various pairs (a, b) and keep those pairs for which $\rho_{(a,b)} \in \mathcal{C}$, ie, $\rho_{(a,b)}(t, X) \in [-1, 1]$. Among those, one can further choose the ones that fulfill some of the desirable properties \mathcal{P} .

REMARK 10.4 The naive numerical simulation of unconditional expectations involving indicators of the type $1_{S_t > K}$ performs well (see Guyon and Henry-Labordère 2014, Table 11.2), except when K is very large or very small. Indeed, in those limiting cases, the ratios of the type

$$P_{0t} \frac{\mathbb{E}^{\mathbb{Q}^t}[(r_t - r_t^0)1_{S_t > K}]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t, K)}$$

become of the indefinite form $\frac{0}{0}$. To circumvent this numerical difficulty, Guyon and Henry-Labordère (2011b) have derived a Malliavin “disintegration by parts” formula that transforms this ratio into a conditional expectation:

$$P_{0t} \frac{\mathbb{E}^{\mathbb{Q}^t}[(r_t - r_t^0)1_{S_t > K}]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t, K)} = \frac{2}{K} \int_0^t \mathbb{E}^{\mathbb{Q}^t}[\sigma_r^t(s) D_s^{B^t} S_t \mid S_t = K] ds.$$

This absorbs the singularity (see Guyon and Henry-Labordère 2014, Section 11.7.6; Guyon and Henry-Labordère 2011b). As a result, the calibrated local volatilities/local correlations are much smoother in the wings (small/large K).

REMARK 10.5 One may wish to correlate the extra Brownian motions that drive the dynamics of the stochastic volatilities and the stochastic interest rates as well as the two Brownian motions that drive the dynamics of the two FX rates. Here is one way to adapt the above method. Assume, for simplicity, that each of the extra processes $a_t^1, a_t^2, r_t^d, r_t^1$ and r_t^2 is driven by exactly one extra Brownian motion. Pick a set C^* of admissible constant values for the twelve correlations:

$$\begin{aligned} C(\rho) = \{&\rho_{S^1 a^1}, \rho_{S^1 r^d}, \rho_{S^1 r^1}, \rho_{a^1 r^d}, \rho_{a^1 r^1}, \\ &\rho_{r^d r^1}, \rho_{S^2 a^2}, \rho_{S^2 r^d}, \rho_{S^2 r^2}, \rho_{a^2 r^d}, \rho_{a^2 r^2}, \rho_{r^d r^2}\}. \end{aligned}$$

The first six correlations are used to calibrate σ_1 ; the last six are used to calibrate σ_2 . Then, one builds two full correlation matrixes ρ^0 and ρ^1 , as follows. First, one picks constant values for all the unspecified correlations in the matrix except $\rho_{S^1 S^2}$. Those values can be arbitrary or inferred from historical data, and they may make the matrix fail to be PSD. Then, one chooses the extremal value $\rho_{S^1 S^2} = -1$ (respectively 1) and projects the resulting matrix onto the space of correlation matrixes to get ρ^0 (respectively ρ^1). The projection method must leave $C(\rho)$ unchanged. This can be done by using weighted norms on matrixes (see, for example, Higham 2002). Then, one assumes that the entire 7×7 correlation matrix $\rho(t, X)$ lies on the line defined by ρ^0 and ρ^1 , $\rho(t, X) = (1 - \lambda(t, X))\rho^0 + \lambda(t, X)\rho^1$; picks two functions $a(t, X)$ and $b(t, X)$ (with b nonvanishing); and, using the particle method, builds $\lambda_{(a,b)}(t, X)$ such that $a + b\lambda_{(a,b)}$ is local in cross and the calibration condition (10.2) is satisfied, with $\rho(t, X_t)$ replaced by $(1 - \lambda_{(a,b)}(t, X_t))\rho_{12}^0 + \lambda_{(a,b)}(t, X_t)\rho_{12}^1$. Then, one has to

verify that $\lambda_{(a,b)}$ takes values in $[0, 1]$. Actually, any ρ^0 and ρ^1 for which $C(\rho^0) = C(\rho^1) = C^*$ will do the job. This guarantees that the knowledge of $\rho(t, X)$ is not needed during the first step of the calibration procedure, ie, the calibration of σ_1 and σ_2 , so that, indeed, the calibration procedure can be cut into two consecutive steps. It is natural to calibrate σ_1 and σ_2 independently of “cross-correlations” such as $\rho_{S^1 S^2}$, $\rho_{S^1 a^2}$, $\rho_{S^2 a^1}$, \dots .

10.2 The equity index smile calibration problem

Let us consider a model that combines local stochastic volatility, stochastic interest rate, stochastic repo (inclusive of the dividend yield) and local correlation. Here, again, let us assume for simplicity that the extra Brownian motions that drive the dynamics of the stochastic volatilities, the stochastic interest rate and the stochastic repos are independent of the Brownian motions (W^1, \dots, W^N) that drive the dynamics of the N stocks (S^1, \dots, S^N) . The correlation matrix of the extra Brownian motions is assumed to be known and constant. Only the correlation of (W^1, \dots, W^N) is unknown; it is assumed to be local:

$$dS_t^i = (r_t - q_t^i)S_t^i dt + \sigma_i(t, S_t^i)a_t^i S_t^i dW_t^i, \quad d\langle W^i, W^j \rangle_t = \rho_{ij}(t, X_t) dt, \quad (10.6)$$

where r_t , q_t^i , a_t^i are stochastic processes, $X_t = (S_t^1, \dots, S_t^N, a_t^1, \dots, a_t^N, D_{0t})$, and $D_{0t} = \exp(-\int_0^t r_s ds)$.³

First, the local volatilities $\sigma_i(t, S^i)$ are calibrated to the market smiles of the S^i using Proposition 22 in the online appendix:

$$\begin{aligned} \sigma_i(t, K)^2 &\frac{\mathbb{E}[D_{0t}(a_t^i)^2 | S_t^i = K]}{\mathbb{E}[D_{0t} | S_t^i = K]} \\ &= \sigma_{\text{Dup}}^i(t, K)^2 - \frac{\mathbb{E}[D_{0t}(r_t - q_t^i - (r_t^0 - q_t^{i,0}))1_{S_t^i > K}]}{\frac{1}{2}K\partial_K^2 \mathcal{C}_i(t, K)} \\ &\quad + \frac{\mathbb{E}[D_{0t}(q_t^i - q_t^{i,0})(S_t^i - K)^+]}{\frac{1}{2}K^2\partial_K^2 \mathcal{C}_i(t, K)}, \end{aligned}$$

where $r_t^0 = -\partial_t \ln P_{0t}$, $q_t^{i,0} = r_t^0 - \partial_t \ln(f_0^{i,t}/S_0^i)$, with $f_0^{i,t}$ the forward of maturity t , and

$$\sigma_{\text{Dup}}^i(t, K)^2 = \frac{\partial_t \mathcal{C}_i(t, K) + (r_t^0 - q_t^{i,0})K\partial_K \mathcal{C}_i(t, K) + q_t^{i,0}\mathcal{C}_i(t, K)}{\frac{1}{2}K^2\partial_K^2 \mathcal{C}_i(t, K)},$$

where $\mathcal{C}_i(t, K)$ is the market price of the call option on S^i , with strike K and maturity t . This is achieved in practice thanks to the particle algorithm (Guyon

³ X_t may actually include any \mathcal{F}_t -measurable random variable (see Remark 3.4).

and Henry-Labordère 2011b, 2014). At this step, the local correlation $\rho(t, X)$ does not need to be known. Then, $\rho(t, X)$ is calibrated to the market smile of the index $I_t = \sum_{i=1}^N \alpha_i S_t^i$ by using the following (see proof in the online appendix).

PROPOSITION 10.6 *The local volatilities σ_i being fixed, model (10.6) is calibrated to the market smile of the index if and only if*

$$\begin{aligned} & \frac{\mathbb{E}_\rho[D_{0t} v_\rho(t, X_t) | I_t = K]}{\mathbb{E}_\rho[D_{0t} | I_t = K]} \\ &= K^2 \sigma_{\text{Dup}}^I(t, K)^2 - K \frac{\mathbb{E}_\rho[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))1_{I_t > K}]}{\frac{1}{2} \partial_K^2 \mathcal{C}(t, K)} \\ &+ \frac{\mathbb{E}_\rho[D_{0t}(q_t - q_t^0)(I_t - K)^+]}{\frac{1}{2} \partial_K^2 \mathcal{C}(t, K)} \end{aligned} \quad (10.7)$$

for all (t, K) , where

$$\begin{aligned} v_\rho(t, X_t) &= \sum_{i,j=1}^N \alpha_i \alpha_j \rho_{ij}(t, X_t) \sigma_i(t, S_t^i) a_t^i \sigma_j(t, S_t^j) a_t^j S_t^i S_t^j, \\ q_t &= \frac{\sum_{i=1}^N \alpha_i S_t^i q_t^i}{\sum_{i=1}^N \alpha_i S_t^i}; \end{aligned} \quad (10.8)$$

r_t^0 and q_t^0 are deterministic interest rate and repo, and

$$\sigma_{\text{Dup}}^I(t, K)^2 = \frac{\partial_t \mathcal{C}(t, K) + (r_t^0 - q_t^0) K \partial_K \mathcal{C}(t, K) + q_t^0 \mathcal{C}(t, K)}{\frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t, K)},$$

where $\mathcal{C}(t, K)$ is the market price of the call option on I , with strike K and maturity t .

Along the same lines as Section 8, assume that the correlation matrix lies on the line defined by two correlation matrixes ρ_0 and ρ_1 , which may depend on (t, X_t) ,

$$\rho(t, X_t) = (1 - \lambda(t, X_t))\rho^0(t, X_t) + \lambda(t, X_t)\rho^1(t, X_t), \quad \lambda(t, X_t) \in \mathbb{R}. \quad (10.9)$$

Then, (10.7) reads

$$\begin{aligned} & \frac{\mathbb{E}_\rho[D_{0t}(v_{\rho^0}(t, X_t) + (v_{\rho^1} - v_{\rho^0})(t, X_t)\lambda(t, X_t)) | I_t = K]}{\mathbb{E}_\rho[D_{0t} | I_t = K]} \\ &= K^2 \sigma_{\text{Dup}}^I(t, K)^2 - K \frac{\mathbb{E}_\rho[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))1_{I_t > K}]}{\frac{1}{2} \partial_K^2 \mathcal{C}(t, K)} \\ &+ \frac{\mathbb{E}_\rho[D_{0t}(q_t - q_t^0)(I_t - K)^+]}{\frac{1}{2} \partial_K^2 \mathcal{C}(t, K)}. \end{aligned}$$

Using two functions $a(t, X)$ and $b(t, X)$ such that b does not vanish and $a + b\lambda$ is local in index, we get the following.

PROPOSITION 10.7 Assume that $\rho(t, X)$, as defined by (10.9), takes values in \mathcal{C} . Then, it is an admissible correlation if and only if there exist two functions a and b such that b does not vanish and λ satisfies the self-consistency equation $\lambda(t, X) = \lambda_{(a,b)}(t, X) \equiv (f(t, I) - a(t, X))/b(t, X)$, with $f(t, I) \equiv N_f(t, I)/D_f(t, I)$ defined by

$$\begin{aligned} N_f(t, K) &= K^2 \sigma_{\text{Dup}}^I(t, K)^2 - K \frac{\mathbb{E}_{\rho(a,b)}[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))1_{I_t > K}]}{\frac{1}{2}\partial_K^2 \mathcal{C}(t, K)} \\ &\quad + \frac{\mathbb{E}_{\rho(a,b)}[D_{0t}(q_t - q_t^0)(I_t - K)^+]}{\frac{1}{2}\partial_K^2 \mathcal{C}(t, K)} \\ &\quad - \frac{1}{\mathbb{E}_{\rho(a,b)}[D_{0t} \mid I_t = K]} \\ &\quad \times \left(\mathbb{E}_{\rho(a,b)} \left[D_{0t} \left(v_{\rho^0}(t, X_t) - \frac{a(t, X_t)}{b(t, X_t)}(v_{\rho^1}(t, X_t) \right. \right. \right. \\ &\quad \left. \left. \left. - v_{\rho^0}(t, X_t)) \mid I_t = K \right] \right), \\ D_f(t, K) &= \frac{\mathbb{E}_{\rho(a,b)}[D_{0t}((v_{\rho^1}(t, X_t) - v_{\rho^0}(t, X_t))/b(t, X_t)) \mid I_t = K]}{\mathbb{E}_{\rho(a,b)}[D_{0t} \mid I_t = K]}, \end{aligned}$$

where $\rho_{(a,b)} = (1 - \lambda_{(a,b)})\rho^0 + \lambda_{(a,b)}\rho^1$.

One can then compute $\rho_{(a,b)}$ using the particle method. Eventually, one has to verify that $\rho_{(a,b)}(t, X)$ is a true correlation matrix. If this is not the case, one may “cap” and “floor” $\rho_{(a,b)}$ (at the edges of $(\rho^0, \rho^1) \cap \mathcal{C}$), when needed, and check how large the resulting smile calibration error is. It is very easy to adapt Remark 10.5 to extend to cases in which the extra Brownian motions are correlated with (W^1, \dots, W^N) .

11 CONCLUSION

Only two local correlation models have been proposed in the past in order to exactly calibrate to the smile of a basket, be it a stock index, a cross-FX rate, an interest rate spread, etc. Both models may actually fail to calibrate the basket smile, and, even if they do not, they impose a particular shape of the correlation matrix that one has no reason to undergo. In this paper, we combined the particle method with a new, simple idea to produce a whole family of local correlation models that calibrate to a basket smile. Our new technique spans all the admissible correlations that belong to a given line (ρ^0, ρ^1) . We have also shown how to build admissible models that combine stochastic interest rates, stochastic dividend yield, local stochastic volatility and local correlation.

The huge number of degrees of freedom, represented by the two functions a and b , allows one to pick one's favorite correlation, with desirable properties, from among the new family of admissible correlations. This way, we reconcile static calibration, ie, calibration from snapshots of prices of options on baskets, and dynamic calibration, ie, calibration from historical study of state-dependency of correlation. Our numerical tests show the wide variety of admissible correlations and give insight into lower bounds/upper bounds on general multi-asset option prices, given the smile of a basket and the smiles of its constituents.

A natural extension of the local correlation model studied in this paper is the cross-dependent volatility model (see Guyon 2016), where the asset volatilities (not only the correlation) depend on the N asset value S_t^1, \dots, S_t^N . Giving up the strong assumption of local volatilities relaxes the constraint on correlation, shifts the focus from correlation to covariance (which is the true quantity of interest) and widens the range of option prices in calibrated models.

Three examples of important open questions that we leave for future work are the derivation of the exact bounds; the derivation of conditions under which a triangle of surfaces of FX-implied volatilities is jointly arbitrage-free; and, when so, the derivation of conditions under which an admissible local correlation does exist in theory.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

I would like to thank Bruno Dupire, Sylvain Corlay, Pierre Henry-Labordère, Pascal Delanoe, Alexey Polishchuk, Fabio Mercurio and Oleg Kovrizhkin for fruitful discussions. I am grateful to Oleg Kovrizhkin and Sylvain Corlay for providing me with the data, to Ivailo Dimov for producing Figure 16, and to the two anonymous referees whose comments helped improve this paper.

REFERENCES

- Ahdida, A., and Alfonsi, A. (2013). A mean-reverting SDE on correlation matrixes. *Stochastic Processes and Their Applications* **123**(4), 1472–1520.
- Austing, P. (2011). Repricing the cross smile: an analytic joint density. *Risk* **24**(7), 72–75 (<http://doi.org/bsrk>).
- Avellaneda, M., Boyer-Olson, D., Busca, J., and Friz, P. (2002). Reconstructing volatility. *Risk* **15**(10), 91–95.

- Bakshi, G., Kapadia, N., and Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies* **16**, 101–143 (<http://doi.org/bpcb>).
- Bollen, N., and Whaley, R. (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance* **59**(2), 711–753 (<http://doi.org/d2dmnx>).
- Branger, N., and Schlag, C. (2004). Why is the index smile so steep? *Review of Finance* **8**(1), 109–127 (<http://doi.org/csppgf>).
- Cont, R., and Deguest, R. (2010). Equity correlations implied by index options: estimation and model uncertainty analysis. Working Paper, Social Science Research Network (<http://doi.org/fxqkp3>).
- Corlay, S. (2016). B-spline techniques for volatility modeling. *The Journal of Computational Finance* **19**(3), 97–135 (<http://doi.org/bsrm>).
- Da Fonseca, J., Grasselli, M., and Tebaldi, C. (2008). Option pricing when correlations are stochastic: an analytical framework. *Review of Derivatives Research* **10**, 151–180 (<http://doi.org/dsgsvh>).
- De Col, A., Gnoatto, A., and Grasselli, M. (2013). Smiles all around: FX joint calibration in a multi-Heston model. *Journal of Banking & Finance* **37**(10), 3799–3818 (<http://doi.org/bsrn>).
- Delanoe, P. (2013). Local correlation with local vol and stochastic vol: towards correlation dynamics? Presentation, April 2013, Global Derivatives Conference.
- Dupire, B. (1994). Pricing with a smile. *Risk* **7**(January), 18–20.
- Dupire, B. (1998). A new approach for understanding the impact of volatility on option prices. Presentation, October 30, Risk Conference.
- Durrleman, V., and El Karoui, N. (2008). Coupling smiles. *Quantitative Finance* **8**(6), 573–590 (<http://doi.org/bcpz49>).
- El Karoui, N., Jeanblanc, M., and Shreve, S. (1998). Robustness of the Black and Scholes formula. *Mathematical Finance* **8**(2), 93–126 (<http://doi.org/dkgzgn>).
- Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. Wiley.
- Gourieroux, C., and Sufana, R. (2003). Wishart quadratic term structure models. Working Paper, Social Science Research Network (<http://doi.org/fx93jp>).
- Guyon, J. (2013). A new class of local correlation models. Working Paper, Social Science Research Network. URL: <http://ssrn.com/abstract=2283419>.
- Guyon, J. (2014a). Local correlation families. *Risk* **27**(2), 52–58.
- Guyon, J. (2014b). Path-dependent volatility. *Risk* **27**(10), 1–7.
- Guyon, J. (2016). Cross-dependent volatility. *Risk* **29**(4), 61–65.
- Guyon, J., and Henry-Labordère, P. (2011a). From spot volatilities to implied volatilities. *Risk* **24**(6), 79–84.
- Guyon, J., and Henry-Labordère, P. (2011b). The smile calibration problem solved. Working Paper, Social Science Research Network. URL: <http://ssrn.com/abstract=1885032>. Shorter version, “Being particular about calibration”, published in *Risk* **25**(1), 92–97. Longer version published in *Post-Crisis Quant Finance*. Risk Books (2013).
- Guyon, J., and Henry-Labordère, P. (2014). *Nonlinear Option Pricing*. Chapman & Hall/CRC Financial Mathematics Series.
- Gyöngy, I. (1986). Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields* **71**, 501–516 (<http://doi.org/dhj4ff>).

- Henry-Labordère, P. (2009). *Analysis, Geometry, and Modeling in Finance*. Chapman & Hall/CRC Financial Mathematics Series.
- Henry-Labordère, P. (2011). Automated option pricing: numerical methods. Working Paper, Social Science Research Network. URL: <http://ssrn.com/abstract=1968344>.
- Higham, N. (2002). Computing a nearest symmetric correlation matrix: a problem from finance. *IMA Journal of Numerical Analysis* **22**(3), 329–343 (<http://doi.org/d8gj5f>).
- Jourdain, B., and Sbai, M. (2012). Coupling index and stocks. *Quantitative Finance* **12**(5), 805–818 (<http://doi.org/c9snn2>).
- Kovrizhkin, O. (2012). Local volatility + local correlation multicurrency model. Presentation, Global Derivatives Conference.
- Langnau, A. (2010). A dynamic model for correlation. *Risk* **23**(4), 74–78.
- Reghai, A. (2010). Breaking correlation breaks. *Risk* **23**(10), 90–95.

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net



Research Paper

Error analysis in Fourier methods for option pricing

Fabián Crocce,^{1,2} Juho Häppölä,¹ Jonas Kiessling³ and
Raúl Tempone¹

¹Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), Building 1, 4700 King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia; emails: fcrocce@gmail.com, juho.happola@iki.fi, raul.tempone@kaust.edu.sa

²Facultad de Ingeniería, Universidad de la República, Ave Julio Herrera y Reissig 565, 11200 Montevideo, Uruguay; email: fcrocce@fing.edu.uy

³Independent; email: jonkie@kth.se

(Received February 28, 2015; revised October 24, 2015; accepted December 2, 2015)

ABSTRACT

We provide a bound for the error committed when using a Fourier method to price European options, when the underlying follows an exponential Lévy dynamic. The price of the option is described by a partial integro-differential equation (PIDE). Applying a Fourier transformation to the PIDE yields an ordinary differential equation (ODE) that can be solved analytically in terms of the characteristic exponent of the Lévy process. Then, a numerical inverse Fourier transform allows us to obtain the option price. We present a bound for the error and use this bound to set the parameters for the numerical method. We analyze the properties of the bound and demonstrate the minimization of the bound to select parameters for a numerical Fourier transformation method in order to solve the option price efficiently.

Keywords: European options; Fourier methods; error analysis; trapezoid quadrature; Lévy processes.

Corresponding author: J. Häppölä

Print ISSN 1460-1559 | Online ISSN 1755-2850
Copyright © 2017 Infopro Digital Limited

1 INTRODUCTION

Lévy processes form a rich field within mathematical finance. They allow the modeling of asset prices with possibly discontinuous dynamics. An early, and probably the best-known, model involving a Lévy process is the Merton (1976) model, which generalizes the Black and Scholes (1973) model. More recently, we have seen more complex models allowing for more general dynamics of the asset price. Examples of such models include the Kou (2002) model (see also Dotsis *et al* 2007), the normal inverse Gaussian model (Barndorff-Nielsen 1997; Rydberg 1997), the variance gamma (VG) model (Madan and Seneta 1990; Madan *et al* 1998) and the Carr–Geman–Madan–Yor (CGMY) model (Carr *et al* 2002, 2003). For a good exposition on jump processes in finance, we refer the reader to Cont and Tankov (2004) (see also Eberlein 2001; Raible 2000).

The prices of European options with underlying assets driven by the Lévy process are solutions to partial integro-differential equations (PIDEs) (Almendral and Oosterlee 2005; Briani *et al* 2004; Kiessling and Tempone 2011; Nualart and Schoutens 2001); they generalize the Black–Scholes equation by incorporating a nonlocal integral term to account for the discontinuities in the asset price. This approach has also been extended to cases where the option price features path dependence (see, for example, Boyarchenko and Levendorskii 2002; d’Halluin *et al* 2004; Lord *et al* 2008).

The Lévy–Khintchine formula provides an explicit representation of the characteristic function of a Lévy process (see Tankov 2004). As a consequence, one can derive an exact expression for the Fourier transform of the solution of the relevant PIDE. Using the inverse fast Fourier transform (iFFT) method, one may efficiently compute the option price for a range of asset prices simultaneously. Further, in the case of European call options, one may use the duality property presented by Dupire (1997) and iFFT to efficiently compute option prices for a wide range of strike prices.

Despite the popularity of Fourier methods for option pricing, few works can be found on error analysis and related parameter selection for these methods. A bound for the error not only provides an interval for the precise value of the option, but also suggests a method for selecting the parameters of the numerical method. An important work in this direction is Lee (2004), in which several payoff functions are considered for a rather general set of models, whose characteristic function is assumed to be known. Feng and Linetsky (2008) presents the framework and theoretical approach for the error analysis and establishes polynomial convergence rates for approximations of the option prices. For a more contemporary review on the error committed in various FT-related methods, we refer the reader to Boyarchenko and Levendorskii (2011). That paper extends the classical flat Fourier methods by deforming the integration

contours on the complex plane; it also looks at the discretely monitored barrier options studied in De Innocentis and Levendorskii (2014).

In this work, we present a methodology for studying and bounding the error committed when using FT methods to compute option prices. We also provide a systematic way of choosing the parameters of the numerical method in a way that minimizes the strict error bound, thus guaranteeing adherence to a pre-described error tolerance. We focus on exponential Lévy processes that may be either diffusive or pure jump. Our contribution is to derive a strict error bound for a Fourier transform method when pricing options under risk-neutral Lévy dynamics. We derive a simplified bound that separates the contributions of the payoff and the process in an easily processed and extensible product form; this is independent of the asymptotic behavior of the option price at extreme prices and strike parameters. We also provide a proof for the existence of optimal parameters of the numerical computation that minimize the presented error bound. When comparing our work with that of Lee (2004), we find that Lee's work is more general than ours, in that he studies a wider range of processes. However, our results apply to a larger class of payoffs. In test examples of practical relevance, we also find that the bound presented produces comparable or better results than those previously presented in the literature, with an acceptable computational cost.

This paper is organized as follows. Section 2 introduces the PIDE setting in the context of risk-neutral asset pricing; we show the Fourier representation of the relevant PIDE for asset pricing with Lévy processes and use that representation for derivative pricing. In Section 3, we derive a representation for the numerical error and divide it into quadrature and cutoff contributions. We also describe the methodology for choosing numerical parameters to obtain minimal error bounds for the FT method. The derivation is supported by numerical examples using relevant test cases with both diffusive and pure-jump Lévy processes in Section 4. Numerics are followed by conclusions in Section 5.

2 FOURIER METHOD FOR OPTION PRICING

Consider an asset whose price at time t is modeled by the stochastic process $S = (S_t)$, defined by $S_t = S_0 e^{X_t}$, where $X = (X_t) \in \mathbb{R}$ is assumed to be a Lévy process whose jump measure ν satisfies

$$\int_{\mathbb{R} \setminus \{0\}} \min\{y^2, 1\} \nu(dy) < \infty. \quad (2.1)$$

Assuming the risk-neutral dynamic for S_t , the price at time $t = T - \tau$ of a European option with payoff G and maturity time T is given by

$$\Pi(\tau, s) = e^{-r\tau} E(G(S_T) | S_{T-\tau} = s),$$

where r is the short rate that we assume to be constant and $\tau : 0 \leq \tau \leq T$ is the time to maturity. Extensions to nonconstant deterministic short rates are straightforward.

The infinitesimal generator of a Lévy process X is given by (see Applebaum 2004)

$$\begin{aligned}\mathcal{L}^X f(x) &\equiv \lim_{h \rightarrow 0} \frac{E(f(X_{t+h}) | X_t = x) - f(x)}{h} \\ &= \gamma f'(x) + \frac{\sigma^2}{2} f''(x) \\ &\quad + \int_{\mathbb{R} \setminus \{0\}} (f(x+y) - f(x) - y \mathbf{1}_{|y| \leq 1} f'(x)) v(dy),\end{aligned}\quad (2.2)$$

where (γ, σ^2, v) is the characteristic triple of the Lévy process. The risk-neutral assumption on (S_t) implies

$$\int_{|y|>1} e^y v(dy) < \infty \quad (2.3)$$

and fixes the drift term (see Kiesel and Tempone 2011) γ of the Lévy process to

$$\gamma = r - \frac{\sigma^2}{2} - \int_{\mathbb{R} \setminus \{0\}} (e^y - 1 - y \mathbf{1}_{|y| \leq 1}) v(dy). \quad (2.4)$$

Thus, the infinitesimal generator of X may be written under the risk-neutral assumption as

$$\begin{aligned}\mathcal{L}^X f(x) &= \left(r - \frac{\sigma^2}{2}\right) f'(x) + \frac{\sigma^2}{2} f''(x) \\ &\quad + \int_{\mathbb{R} \setminus \{0\}} (f(x+y) - f(x) - (e^y - 1) f'(x)) v(dy).\end{aligned}\quad (2.5)$$

Consider g as the reward function in log prices (ie, defined by $g(x) = G(S_0 e^x)$). Now, take f to be defined as

$$f(\tau, x) \equiv E(g(X_T) | X_{T-\tau} = x).$$

Then, f solves the following PIDE:

$$\begin{aligned}\partial_\tau f(\tau, x) &= \mathcal{L}^X f(\tau, x), \\ f(0, x) &= g(x), \quad (\tau, x) \in [0, T] \times \mathbb{R}.\end{aligned}$$

Observe that f and Π are related by

$$\Pi(\tau, S_0 e^x) = e^{-r\tau} f(\tau, x). \quad (2.6)$$

Consider a damped version of f defined by $f_\alpha(\tau, x) = e^{-\alpha x} f(\tau, x)$; we see that $\partial_\tau f_\alpha = e^{-\alpha x} \mathcal{L}^X f(\tau, x)$.

There are different conventions for the Fourier transform. Here, we consider the operator \mathcal{F} such that

$$\mathcal{F}[f](\omega) \equiv \int_{\mathbb{R}} e^{i\omega x} f(x) dx, \quad (2.7)$$

defined for functions f for which the previous integral is convergent. We also use $\hat{f}(\omega)$ as a shorthand notation of $\mathcal{F}[f](\omega)$. To recover the original function f , we define the inverse Fourier transform as

$$\mathcal{F}^{-1}[f](x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\omega x} f(\omega) d\omega.$$

We have that $\mathcal{F}^{-1}[\hat{f}](x) = f(x)$.

Applying \mathcal{F} to f_α , we get $\hat{f}_\alpha(\omega) = \hat{f}(\omega + i\alpha)$. Observe also that the Fourier transform applied to $\mathcal{L}^X f(\tau, x)$ gives $\Psi(-i\omega) \hat{f}(\tau, \omega)$, where $\Psi(\cdot)$ is the characteristic exponent of the process X , which satisfies $E(e^{zX_t}) = e^{t\Psi(z)}$. The explicit expression for $\Psi(\cdot)$ is

$$\Psi(z) = \left(r - \frac{\sigma^2}{2} \right) z + \frac{\sigma^2}{2} z^2 + \int_{\mathbb{R}} (e^{zy} - 1 - (e^y - 1)z) \nu(dy). \quad (2.8)$$

From the previous considerations, it can be concluded that

$$\partial_\tau \hat{f}_\alpha = \Psi(\alpha - i\omega) \hat{f}(\omega - i\alpha). \quad (2.9)$$

Now, $\hat{f}(\omega - i\alpha) = \hat{f}_\alpha(\omega)$, so \hat{f}_α satisfies the following ODE:

$$\begin{aligned} \frac{\partial_\tau \hat{f}_\alpha(\tau, \omega)}{\hat{f}_\alpha(\tau, \omega)} &= \Psi(\alpha - i\omega), \\ \hat{f}_\alpha(0, \omega) &= \hat{g}_\alpha(\omega). \end{aligned} \quad (2.10)$$

Solving the previous ODE explicitly, we obtain

$$\hat{f}_\alpha(\tau, \omega) = e^{\tau\Psi(\alpha - i\omega)} \hat{g}_\alpha(\omega). \quad (2.11)$$

Observe that the first factor on the right-hand side of the above equation is $E(e^{(\alpha - i\omega)X_\tau})$ (ie, $\varphi_\tau(-i\alpha - \omega)$), where $\varphi_\tau(\cdot)$ denotes the characteristic function of the random variable X_τ :

$$\varphi_\tau(\omega) \equiv E(\tau\Psi(i\omega)). \quad (2.12)$$

Now, we employ the inverse Fourier transformation to obtain the value function:

$$f_\alpha(\tau, x) = \mathcal{F}^{-1}[\hat{f}_\alpha](\tau, x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\omega x} \hat{f}_\alpha(\tau, \omega) d\omega, \quad (2.13)$$

or

$$f_\alpha(\tau, x) = \frac{1}{\pi} \int_0^{+\infty} \operatorname{Re}[e^{-i\omega x} \hat{f}_\alpha(\tau, \omega)] d\omega. \quad (2.14)$$

As it is typically not possible to compute the inverse Fourier transform analytically, we approximate it by discretizing and truncating the integration domain using trapezoidal quadrature (2.13). Consider the following approximation:

$$f_{\alpha, \Delta\omega, n}(\tau, x) = \frac{\Delta\omega}{2\pi} \sum_{k=-n}^{n-1} e^{-i(k+\frac{1}{2})\Delta\omega x} \hat{f}_\alpha(\tau, (k + \frac{1}{2})\Delta\omega) \quad (2.15)$$

$$= \frac{\Delta\omega}{\pi} \sum_{k=0}^{n-1} \operatorname{Re}[e^{-i(k+\frac{1}{2})\Delta\omega x} \hat{f}_\alpha(\tau, (k + \frac{1}{2})\Delta\omega)]. \quad (2.16)$$

Bounding and consequently minimizing the error in the approximation of $f(\tau, x)$ by

$$f_{\Delta\omega, n}(\tau, x) \equiv e^{\alpha x} f_{\alpha, \Delta\omega, n}(\tau, x)$$

is the main focus of this paper and will be addressed in the following section.

REMARK 2.1 Although we are mainly concerned with option pricing when the payoff function can be damped in order to guarantee regularity in the L^1 sense, we note here that our main results are naturally extendable to include the Greeks of the option. Indeed, we have by (2.11) that

$$f(t, x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{(\alpha - i\omega)x} \hat{f}_\alpha(\tau, \omega) d\omega, \quad (2.17)$$

so the Delta and Gamma of the option equal

$$\Delta(t, x) \equiv \frac{\partial f(t, x)}{\partial x} = \frac{1}{2\pi} \int_{\mathbb{R}} (\alpha - i\omega) e^{(\alpha - i\omega)x} \hat{f}_\alpha(\tau, \omega) d\omega, \quad (2.18)$$

$$\Gamma(t, x) \equiv \frac{\partial^2 f(t, x)}{\partial x^2} = \frac{1}{2\pi} \int_{\mathbb{R}} (\alpha - i\omega)^2 e^{(\alpha - i\omega)x} \hat{f}_\alpha(\tau, \omega) d\omega. \quad (2.19)$$

Because the expressions involve partial derivatives with respect to only x , the results in this work are applicable to the computation of Δ and Γ through a modification of the payoff function:

$$\hat{g}_{\alpha, \Delta}(\omega) = \hat{g}_\alpha(\omega)(\alpha - i\omega), \quad (2.20)$$

$$\hat{g}_{\alpha, \Gamma}(\omega) = \hat{g}_\alpha(\omega)(\alpha - i\omega)^2. \quad (2.21)$$

When the Fourier space payoff function manifests exponential decay, the introduction of a coefficient that is polynomial in ω does not change the regularity of \hat{g} in a way

that would significantly change the following analysis. Last, we note that since we do our analysis for PIDEs on a mesh of x 's, we may also compute the option values in one go and obtain the Greeks with little additional effort, using a finite difference approach for the derivatives.

2.1 Evaluation of the method for multiple values of x simultaneously

The fast Fourier transform (FFT) algorithm provides an efficient way of computing (2.15) for an equidistantly spaced mesh of values for x simultaneously. Examples of works that consider this widely extended tool are Lord *et al* (2008), Jackson *et al* (2008), Hurd and Zhou (2010) and Schmelzle (2010).

Similarly, one may define the Fourier frequency ω as the conjugate variable of some external parameter on which the payoff depends. In particular, for the practically relevant case of call options, we can denote the log-strike as k , treat x as a constant and write

$$\bar{f}_{k,\alpha}(\omega) \equiv \int_{\mathbb{R}} e^{(\alpha+i\omega)k} f_k(x) dk. \quad (2.22)$$

Using this convention, the time dependence is given by

$$\tilde{f}_{k,\alpha}(\tau, x) = \frac{e^{(i\omega+\alpha+1)x} \varphi_\tau(\omega - i(\alpha + 1))}{(i\omega + \alpha)(i\omega + \alpha + 1)} \quad (2.23)$$

contrasted with the x -space solution

$$\hat{f}_{k,\alpha}(\tau, x) = \frac{e^{(i\omega-\alpha+1)k} \varphi_\tau(\omega + i\alpha)}{(i\omega + \alpha)(i\omega + \alpha + 1)}. \quad (2.24)$$

We note that for a call option payoff to be in L^1 , we demand that α in (2.23) is positive. By omitting the exponential factors that contain the x and k dependence in (2.23) and (2.24), respectively, one can get from (2.23) to (2.24) using the mapping $\alpha \mapsto -\alpha - 1$. Thanks to this, much of the analysis regarding the x -space transformation generalizes in a straightforward manner to the k -space transform.

3 ERROR BOUND

The aim of this section is to compute a bound of the error when approximating the option price $f(\tau, x)$ by $f_{\alpha, \Delta\omega, n}(\tau, x)$, which is defined in (2.15). Considering

$$f_{\alpha, \Delta\omega}(\tau, x) = \frac{\Delta\omega}{2\pi} \sum_{k \in \mathbb{Z}} e^{-i(k + \frac{1}{2})\Delta\omega x} \hat{f}_\alpha(\tau, (k + \frac{1}{2})\Delta\omega), \quad (3.1)$$

the total error \mathcal{E} can be split into a sum of two terms: the quadrature and truncation errors. The former is the error from the approximation of the integral in (2.13) by the

infinite sum in (3.1), while the latter is due to the truncation of the infinite sum. Using the triangle inequality, we have

$$\mathcal{E} := |f(\tau, x) - f_{\Delta\omega, n}(\tau, x)| \leq \mathcal{E}_{\mathcal{Q}} + \mathcal{E}_{\mathcal{F}} \quad (3.2)$$

with

$$\begin{aligned}\mathcal{E}_{\mathcal{Q}} &= e^{\alpha x} |f_{\alpha}(\tau, x) - f_{\alpha, \Delta\omega}(\tau, x)|, \\ \mathcal{E}_{\mathcal{F}} &= e^{\alpha x} |f_{\alpha, \Delta\omega}(\tau, x) - f_{\alpha, \Delta\omega, n}(\tau, x)|.\end{aligned}$$

Observe that each \mathcal{E} , $\mathcal{E}_{\mathcal{Q}}$ and $\mathcal{E}_{\mathcal{F}}$ depends on three kinds of parameters:

- those underlying the model and payoff, such as volatility and strike price, which we call physical parameters;
- those relating to the numerical scheme, such as α and n ;
- auxiliary parameters, which will be introduced in the process of deriving the error bound; these parameters do not enter the computation of the option price, but they need to be chosen appropriately to have as tight a bound as possible.

We start by analyzing the quadrature error.

3.1 Quadrature error

Denote by A_a , with $a > 0$, the strip of width $2a$ around the real line

$$A_a \equiv \{z \in \mathbb{C}: |\text{Im}[z]| < a\}.$$

The following theorem presents conditions under which the quadrature error goes to zero at a spectral rate as $\Delta\omega$ goes to zero. Later in this section, we will discuss simpler conditions to verify the hypotheses and analyze in more detail the case where the process X is a diffusive process, or there are “enough small jumps”.

THEOREM 3.1 *Assume that, for $a > 0$,*

(H1) *the characteristic function of the random variable X_1 has an analytic extension to the set*

$$A_a - \alpha i \equiv \{z \in \mathbb{C}: |\text{Im}[z] + \alpha| < a\},$$

(H2) *the Fourier transform of $g_{\alpha}(x)$ is analytic in the strip A_a ,*

(H3) *there exists a continuous function $\gamma \in L^1(\mathbb{R})$ such that $|\hat{f}_{\alpha}(\tau, \omega + i\beta)| < \gamma(\omega)$ for all $\omega \in \mathbb{R}$ and for all $\beta \in [-a, a]$.*

Then, the quadrature error is bounded by

$$\mathcal{E}_Q \leq e^{\alpha x} \frac{M_{\alpha,a}(\tau, x)}{2\pi(e^{2\pi a/\Delta\omega} - 1)},$$

where $M_{\alpha,a}(\tau, x)$ is given by

$$M_{\alpha,a}(\tau, x) := \sum_{\beta \in \{-a, a\}} \int_{\mathbb{R}} |e^{-i(\omega+i\beta)x} \hat{f}_{\alpha}(\tau, \omega + i\beta)| d\omega. \quad (3.3)$$

$M_{\alpha,a}(\tau, x)$ equals the Hardy norm (defined in (3.4)) of the function

$$\omega \mapsto e^{-i(\omega+i\beta)x} \hat{f}_{\alpha}(\tau, \omega + i\beta),$$

which is finite.

The proof of Theorem 3.1 is an application of Stenger (1993, Theorem 3.2.1), whose relevant parts we include for ease of reading. Using the notation in Stenger (1993), $H_{A_a}^1$ is the family of functions w that are analytic in A_a , such that

$$\|w\|_{H_{A_a}^1} := \lim_{\varepsilon \rightarrow 0} \int_{\partial A_a(\varepsilon)} |w(z)| d|z| < \infty, \quad (3.4)$$

where

$$A_a(\varepsilon) = \left\{ z \in \mathbb{C} : |\operatorname{Re}[z]| < \frac{1}{\varepsilon}, |\operatorname{Im}[z]| < a(1 - \varepsilon) \right\}.$$

LEMMA 3.2 (Stenger 1993, Theorem 3.2.1) *Let $w \in H_{A_a}^1$; then, define*

$$I(w) = \int_{\mathbb{R}} w(x) dx, \quad (3.5)$$

$$J(w, h) = h \sum_{j=-N}^N w(jh), \quad (3.6)$$

$$\xi(w, h) = I(w) - J(w, h), \quad (3.7)$$

and then

$$|\xi(w, h)| \leq \frac{e^{-\pi a/h} \|w\|_{H_{A_a}^1}}{2\sinh(\pi a/h)}. \quad (3.8)$$

PROOF OF THEOREM 3.1 First, observe that H1 and H2 imply that the function $w(z) = e^{-izx+\Delta\omega/2} \hat{f}_{\alpha}(\tau, z + \Delta\omega/2)$ is analytic in A_a . H3 allows us to use dominated convergence theorem to prove that $\|w\|_{H_{A_a}^1}$ is finite and coincides with $M_{\alpha,a}(\tau, x)$. Applying Lemma 3.2, the proof is completed. \square

Regarding the hypotheses of Theorem 3.1, the next propositions provide simpler conditions that imply H1 and H2, respectively.

PROPOSITION 3.3 *If α , a and v are such that*

$$\int_{y>1} e^{(\alpha+a)y} v(dy) < \infty \quad \text{and} \quad \int_{y<-1} e^{(\alpha-a)y} v(dy) < \infty, \quad (3.9)$$

and then H1 in Theorem 3.1 is fulfilled.

PROOF Denoting by $\varphi_1(\cdot)$ the characteristic function of X_1 , we want to prove that $z \mapsto \varphi_1(z + \alpha i)$ is analytic in A_a . Considering that $\varphi_1(z + \alpha i) = e^{\psi(iz - \alpha)}$, the only nontrivial part of the proof is to verify that

$$z \mapsto \int p(z, y) v(dy) \quad (3.10)$$

is analytic in A_a , where $p: A_a \times \mathbb{R} \rightarrow \mathbb{C}$ is given by

$$p(z, y) = e^{y(iz - \alpha)} - 1 - (e^y - 1)(iz - \alpha).$$

To prove this fact, we demonstrate that we can apply the main result and the only theorem in Mattner (2001), which, given a measure space $(\Omega, \mathcal{A}, \mu)$ and an open subset $G \subseteq \mathbb{C}$, ensures the analyticity of $\int f(\cdot, \omega) d\mu(\omega)$, provided that $f: G \times \Omega \rightarrow \mathbb{C}$ satisfies the following: $f(z, \cdot)$ is \mathcal{A} -measurable for all $z \in G$, $f(\cdot, \omega)$ is holomorphic for all $\omega \in \Omega$ and $\int |f(\cdot, \omega)| d\mu(\omega)$ is locally bounded. In our case, we consider the measure space to be \mathbb{R} , with the Borel σ -algebra and the Lebesgue measure, $G = A_a$ and $f = p$. It is clear that $p(x, \cdot)$ is Borel measurable and $p(\cdot, y)$ is holomorphic. It remains for us to verify that

$$z \mapsto \int_{\mathbb{R}^*} |p(z, y)| v(dy)$$

is locally bounded. To this end, we assume that $\operatorname{Re}[z] < b$ (and, since $z \in A_a$, $\operatorname{Im}[z] < a$) and split the integration domain into $|y| > 1$ and $0 < |y| \leq 1$ to prove that both integrals are uniformly bounded.

Regarding the integral in $|y| > 1$, we observe that

$$|p(z, y)| \leq e^{y(\alpha + \operatorname{Im}[z])} + 1 + (e^y + 1)(\alpha + a + b); \quad (3.11)$$

for $y < -1$, we have $e^{y(\alpha + \operatorname{Im}[z])} < e^{y(\alpha - a)}$, while for $y > 1$, we have $e^{y(\alpha + \operatorname{Im}[z])} < e^{y(\alpha + a)}$. Using the previous bounds and hypotheses together with (2.1) and (2.3), we obtain the needed bound.

For the integral in $0 < |y| \leq 1$, observe that, denoting $f(z, y) = |p(z, y)|$, we have $f(z, 0) = 0$ for every z , $\partial_y f(z, 0) = 0$ for every z and $|\partial_{yy} f(z, y)| < c$ for

$z \in A_a$, $\operatorname{Re}[z] < b$, $|y| < 1$. From these observations, we get that the Maclaurin polynomial of degree 1 of $y \mapsto f(z, y)$ is null for every z . We can bound $f(z, y)$ by the remainder term, which, in our region of interest, is bounded by $(c/2)y^2$; thus, we obtain

$$\int_{0 < |y| \leq 1} |p(z, y)|\nu(dy) \leq \frac{c}{2} \int_{0 < |y| \leq 1} y^2 \nu(dy), \quad (3.12)$$

which is finite by the hypothesis on ν . This finishes the proof. \square

PROPOSITION 3.4 *If, for all $b < a$, the function $x \mapsto e^{bx}|g_\alpha(x)$ is in $L^2(\mathbb{R})$, then H2 in Theorem 3.1 is fulfilled.*

PROOF The proof is a direct application of Reed and Simon (1975, Theorem IX.13). \square

We now turn our attention to a more restricted class of Lévy processes; namely, processes such that either $\sigma^2 > 0$ or there exists $\lambda \in (0, 2)$ such that $C(\lambda)$ defined in (3.13) is strictly positive. For this class of processes, we can state our main result explicitly in terms of the characteristic triplet.

Given $\lambda \in (0, 2)$, define $C(\lambda)$ as

$$C(\lambda) = \inf_{\kappa > 1} \left\{ \kappa^\lambda \int_{0 < |y| < 1/\kappa} y^2 \nu(dy) \right\}. \quad (3.13)$$

Observe that $C(\lambda) \geq 0$, and, by our assumptions on the jump measure ν , $C(\lambda)$ is finite. Further, if $\lambda \in (0, 2)$ is such that

$$\liminf_{\epsilon \downarrow 0} \frac{1}{\epsilon^\lambda} \int_{0 < |y| < \epsilon} y^2 \nu(dy) > 0, \quad (3.14)$$

then $C(\lambda) > 0$. To see this, note that (3.14) implies the existence of ϵ_0 , such that

$$\inf_{\epsilon \leq \epsilon_0} \left\{ \frac{1}{\epsilon^\lambda} \int_{0 < |y| < \epsilon} y^2 \nu(dy) \right\} > 0.$$

If $\epsilon_0 < 1$, observe that

$$\inf_{0 \leq \epsilon \leq 1} \left\{ \frac{1}{\epsilon^\lambda} \int_{0 < |y| < \epsilon} y^2 \nu(dy) \right\} \geq \int_{0 < |y| < \epsilon_0} y^2 \nu(dy) > 0,$$

where, for the first inequality, it was taken into account that $1/\epsilon^\lambda \geq 1$ and that the integral is increasing with ϵ . By combining the two previous infima and considering $|\kappa| = 1/\epsilon$, we get that $C(\lambda) > 0$.

Further, we note that for a Lévy model with finite jump intensity, such as the Black–Scholes and Merton models that satisfy the first of our assumptions, $C(\lambda) = 0$ for all $\lambda \in (0, 2)$.

THEOREM 3.5 Assume that α and a are such that (3.9) holds, $\hat{g}_\alpha \in L_{A_a}^\infty$ and either $\sigma^2 > 0$ or $C(\lambda) > 0$ for some $\lambda \in (0, 2)$. Then, the quadrature error is bounded by

$$\varepsilon_Q \leq e^{\alpha x} \frac{\tilde{M}_{\alpha,a}(\tau, x)}{2\pi(e^{2\pi a/\Delta\omega} - 1)},$$

where

$$\begin{aligned} \tilde{M}_{\alpha,a}(\tau, x) &= \sum_{c \in \{-1, 1\}} e^{cax} e^{\tau\Psi(ca)} |\hat{g}_\alpha(ca)| \\ &\times \int_{\mathbb{R}} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \mathbf{1}_{|\omega|>1}\right)\right) d\omega. \end{aligned} \quad (3.15)$$

Further, if $\sigma^2 > 0$, we have

$$\tilde{M}_{\alpha,a}(\tau, x) \leq \frac{\sqrt{2\pi}}{\sigma\sqrt{\tau}} \sum_{c \in \{-1, 1\}} e^{cax} e^{\tau\Psi(ca)} |\hat{g}_\alpha(ca)|. \quad (3.16)$$

PROOF Considering $h_{\alpha,a}(\tau, x, \omega)$, defined by

$$h_{\alpha,a}(\tau, x, \omega) = \sum_{c \in \{-1, 1\}} |e^{-i(\omega+ica)x} \hat{f}_\alpha(\tau, \omega + ica)|, \quad (3.17)$$

we have that

$$M_{\alpha,a}(\tau, x) = \int_{\mathbb{R}} h_{\alpha,a}(\tau, x, \omega) d\omega.$$

However, for $\beta \in (-a, a)$,

$$\begin{aligned} |e^{-i(\omega+i\beta)x} \hat{f}_\alpha(\tau, \omega + i\beta)| &= e^{\beta x} |\hat{f}_\alpha(\tau, \omega + i\beta)| \\ &= e^{\beta x} |e^{\tau\Psi(\alpha+\beta-i\omega)}| |\hat{g}_\alpha(\omega + i\beta)|. \end{aligned} \quad (3.18)$$

For the factor involving the characteristic exponent, we have

$$|e^{\tau\Psi(\alpha+\beta-i\omega)}| = e^{\tau \operatorname{Re}[\Psi(\alpha+\beta-i\omega)]}. \quad (3.19)$$

Now, observe that

$$\begin{aligned} \operatorname{Re}[\Psi(\alpha + \beta - i\omega)] &= (\alpha + \beta) \left(r - \frac{\sigma^2}{2} \right) + \frac{\sigma^2}{2} ((\alpha + \beta)^2 - \omega^2) \\ &+ \int_{\mathbb{R} \setminus \{0\}} (e^{(\alpha+\beta)y} \cos(-y\omega) - 1 - (\alpha + \beta)(e^y - 1)) v(dy). \end{aligned} \quad (3.20)$$

If $|\omega| \leq 1$, we bound $\cos(-y\omega)$ by 1, getting

$$\begin{aligned} \operatorname{Re}[\Psi(\alpha + \beta - i\omega)] &\leq (\alpha + \beta) \left(r - \frac{\sigma^2}{2} \right) + \frac{\sigma^2}{2} ((\alpha + \beta)^2 - \omega^2) \\ &\quad + \int_{\mathbb{R} \setminus \{0\}} (e^{(\alpha+\beta)y} - 1 - (\alpha + \beta)(e^y - 1)) v(dy) \\ &= \Psi(\alpha + \beta) - \frac{\sigma^2}{2} \omega^2. \end{aligned} \quad (3.21)$$

Assume that $|\omega| > 1$. Using that for $|x| < 1$ it holds that $\cos(x) < 1 - x^2/4$, we can bound the first term of the integral in the following manner:

$$\begin{aligned} &\int_{\mathbb{R} \setminus \{0\}} e^{(\alpha+\beta)y} \cos(y\omega) v(dy) \\ &\leq \int_{0 < |y| < 1/|\omega|} e^{(\alpha+\beta)y} (1 - \omega^2 y^2/4) v(dy) + \int_{|y| \geq 1/|\omega|} e^{(\alpha+\beta)y} v(dy) \\ &\leq \int_{\mathbb{R} \setminus \{0\}} e^{(\alpha+\beta)y} v(dy) - \frac{|\omega|^{2-\lambda}}{4} |\omega|^\lambda \int_{0 < |y| < 1/|\omega|} y^2 v(dy) \\ &\leq \int_{\mathbb{R} \setminus \{0\}} e^{(\alpha+\beta)y} v(dy) - \frac{|\omega|^{2-\lambda}}{4} C(\lambda). \end{aligned} \quad (3.22)$$

Inserting (3.22) back into (3.20), we get

$$\begin{aligned} \operatorname{Re}[\Psi(\alpha + \beta - i\omega)] &\leq (\alpha + \beta) \left(r - \frac{\sigma^2}{2} \right) + \frac{\sigma^2}{2} ((\alpha + \beta)^2 - \omega^2) \\ &\quad + \int_{\mathbb{R} \setminus \{0\}} (e^{(\alpha+\beta)y} - 1 - (\alpha + \beta)(e^y - 1)) v(dy) \\ &\quad - \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \\ &= \Psi(\alpha + \beta) - \frac{\sigma^2}{2} \omega^2 - \frac{|\omega|^{2-\lambda}}{4} C(\lambda). \end{aligned}$$

Taking the previous considerations and integrating in \mathbb{R} with respect to ω , we obtain (3.15).

Finally, observing that $C(\lambda) \geq 0$ and bounding it by 0, the bound (3.16) is obtained by evaluating the integral. \square

REMARK 3.6 In the case of call options, hypothesis H2 implies a dependence between the strip-width parameter a and the damping parameter α . We have that the damped payoff of the call option is in $L^1(\mathbb{R})$ if and only if $\alpha > 1$; hence, the appropriate choice of strip-width parameter is given by $0 < a < \alpha - 1$. A similar argument holds for the case of put options, for which the Fourier-transformed damped

payoff is identical to the calls, with the distinction that $\alpha < 0$. In such a case, we require $a < -\alpha$.

In the case of binary options with payoffs that have finite support ($G(x) = \mathbf{1}_{[x_-, x_+]}(x)$), we can set any $a \in \mathbb{R}$ (ie, no damping is needed at all, and even if such damping is chosen, it has no effect on the appropriate choice of a).

REMARK 3.7 The bound we provide for the quadrature error is naturally positive and increasing in $\Delta\omega$. It decays to zero at a spectral rate as $\Delta\omega$ decreases to zero.

3.2 Frequency truncation error

The frequency truncation error is given by

$$\mathcal{E}_F = \frac{e^{\alpha x} \Delta\omega}{\pi} \left| \sum_{k=n}^{\infty} \operatorname{Re}[e^{-i(k+\frac{1}{2})\Delta\omega x} \hat{f}_\alpha(\tau, (k + \frac{1}{2})\Delta\omega)] \right|.$$

If a function $c : (\omega_0, \infty) \rightarrow (0, \infty)$ satisfies

$$|\operatorname{Re}[e^{-i(k+\frac{1}{2})\Delta\omega x} \hat{f}_\alpha(\tau, (k + \frac{1}{2})\Delta\omega)]| \leq c((k + \frac{1}{2})\Delta\omega) \quad (3.23)$$

for every natural number k , then we have that

$$\begin{aligned} \mathcal{E}_F &\leq \frac{e^{\alpha x} \Delta\omega}{\pi} \sum_{k=n}^{\infty} |\operatorname{Re}[e^{-i(k+\frac{1}{2})\Delta\omega x} \hat{f}_\alpha(\tau, (k + \frac{1}{2})\Delta\omega)]| \\ &\leq \frac{e^{\alpha x} \Delta\omega}{\pi} \sum_{k=n}^{\infty} c((k + \frac{1}{2})\Delta\omega). \end{aligned}$$

Further, if c is a nonincreasing concave integrable function, we get

$$\mathcal{E}_F \leq \frac{e^{\alpha x}}{\pi} \int_{n\Delta\omega}^{\infty} c(\omega) d\omega. \quad (3.24)$$

When $\hat{g}_\alpha \in L_{[\omega_0, \infty)}^\infty$ and either $\sigma^2 > 0$ or $C(\lambda) > 0$, then the function c in (3.23) can be chosen as

$$c(\omega) = \|\hat{g}_\alpha\|_{L_{[\omega_0, \infty)}^\infty} e^{\tau\Psi(\alpha)} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \mathbf{1}_{|\omega|>1}\right)\right). \quad (3.25)$$

To prove that this function satisfies (3.23), we can use the same bound we found in the proof of Theorem 3.5, with $\beta = 0$, to obtain

$$\operatorname{Re}[\Psi(\alpha - i\omega)] \leq \Psi(\alpha) - \frac{\sigma^2}{2}\omega^2 - \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \mathbf{1}_{|\omega|>1},$$

from which the result is straightforward.

3.3 Bound for the full error

In this section, we summarize the bounds obtained for the error under different assumptions and analyze their central properties.

In general, the bounds provided in this paper are of the form

$$\bar{\mathcal{E}} = \frac{e^{\alpha x}}{\pi} \left(\frac{\bar{M}}{e^{2\pi a/\Delta\omega} - 1} + \int_{n\Delta\omega}^{\infty} c(\omega) d\omega \right), \quad (3.26)$$

where \bar{M} is an upper bound of $M_{\alpha,a}(\tau, x)$ defined in (3.3), and c is nonincreasing, integrable and satisfies (3.23). Both \bar{M} and c may depend on the parameters of the model and the artificial parameters, but they are independent of $\Delta\omega$ and n . Typically, one can remove the dependence of some of the parameters, simplifying the expressions but obtaining less tight bounds.

When analyzing the behavior of the bound, one can observe that the term correspondent with the quadrature error decreases to zero spectrally when $\Delta\omega$ goes to zero. The second term goes to zero if $n\Delta\omega$ diverges, but we are unable to determine the rate of convergence without further assumptions.

Once an expression for the error bound is obtained, the problem of how to choose the parameters of the numerical method in order to minimize the bound arises, assuming a constraint on the computational effort one is willing to use. The computational effort of the numerical method depends only on n . For this reason, we aim to find the parameters that minimize the bound for a fixed n . The following result shows that the bound obtained, as a function of $\Delta\omega$, has a unique local minimum, which is the global minimum.

PROPOSITION 3.8 *Fix α , a , n and λ , and consider the bound $\bar{\mathcal{E}}$ as a function of $\Delta\omega$. There exists an optimal $\Delta\omega^* \in [\omega_0/n, \infty)$ such that $\bar{\mathcal{E}}$ is decreasing in $(\omega_0/n, \Delta\omega^*)$ and increasing in $(\Delta\omega^*, \infty)$; thus, a global minimum of $\bar{\mathcal{E}}$ is attained at $\Delta\omega^*$.*

Further, the optimal $\Delta\omega$ is either the only point at which $\Delta\omega \mapsto p(n\Delta\omega, b) - c(n\Delta\omega)$, with p defined in (3.27), changes sign, or $\Delta\omega = \omega_0/n$ if $p(\omega_0, b) - c(\omega_0) > 0$.

PROOF Let us simplify the notation by calling $y = n\Delta\omega$, $b = 2\pi a n$ and $\tilde{\mathcal{E}} = \pi e^{-\alpha x} \bar{\mathcal{E}}$. We want to prove the existence of $y^* : y^* \geq \omega_0$ such that $\tilde{\mathcal{E}}(y)$ is decreasing for $\omega_0 < y < y^*$ and increasing for $y > y^*$. We have

$$\tilde{\mathcal{E}}(y) = \frac{\bar{M}}{e^{b/y} - 1} + \int_y^{\infty} c(\omega) d\omega.$$

The first term is differentiable with respect to y and goes to 0 if $y \rightarrow 0^+$. This allows us to express it as an integral of its derivative. We can then express $\tilde{\mathcal{E}}(y)$ as

$$\tilde{\mathcal{E}}(y) = \tilde{\mathcal{E}}(\omega_0) + \int_{\omega_0}^y \left(\frac{b\bar{M}e^{b/\omega}}{(e^{b/\omega} - 1)^2 \omega^2} - c(\omega) \right) d\omega.$$

The first term on the right-hand side of the previous equation is constant. Now, we move on to proving that the integrand is increasing with y and is positive if y is large enough. This can be denoted by

$$p(y, b) = \frac{b\bar{M}e^{b/y}}{(e^{b/y} - 1)^2 y^2}. \quad (3.27)$$

Taking into account that c is integrable, we can compute the limit of the integrand in ∞ , obtaining

$$\lim_{y \rightarrow +\infty} p(y, b) - c(y) = \frac{\bar{M}}{b} > 0.$$

Let us prove that $p(y, b)$ is increasing with y for all $b > 0$, which renders $p(y, b) - c(y)$ also increasing with y . The derivative of p with respect to y is given by

$$\partial_y p(y, b) = \frac{b\bar{M}e^{b/y}((b/y)e^{b/y} - 2e^{b/y} + b/y + 2)}{y^3(e^{b/y} - 1)^3},$$

in which the denominator and the first factor in the numerator are clearly positive. To prove that the remaining factor is also positive, observe that

$$xe^x - 2e^x + x + 2 > 0$$

if $x > 0$. □

3.4 Explicit error bounds

In the case where either $\sigma^2 > 0$ or $C(\lambda) > 0$ for some $\lambda \in (0, 2)$, we can give an explicit version of (3.26). Substituting M with \tilde{M} (defined in Theorem 3.5) and c with the function given in (3.25), we obtain

$$\bar{\varepsilon} = \bar{\varepsilon}_Q + \bar{\varepsilon}_F, \quad (3.28)$$

where

$$\begin{aligned} \bar{\varepsilon}_Q &= \sum_{c \in \{-1, 1\}} \frac{e^{\alpha x} e^{cax} e^{\tau\Psi(ca)} |\hat{g}_\alpha(ca)|}{\pi(e^{(2\pi a/\Delta\omega)} - 1)} \\ &\quad \times \int_{\mathbb{R}} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4}C(\lambda)\mathbf{1}_{|\omega|>1}\right)\right) d\omega, \end{aligned} \quad (3.29)$$

$$\bar{\varepsilon}_F = \frac{e^{\alpha x}}{\pi} \|\hat{g}_\alpha\|_{L^\infty_{\mathbb{R}}} e^{\tau\Psi(\alpha)} \int_{n\Delta\omega}^{\infty} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4}C(\lambda)\mathbf{1}_{|\omega|>1}\right)\right) d\omega. \quad (3.30)$$

This reproduces the essential features of Feng and Linetsky (2008, Theorem 6.6); the bound (3.30) can be further improved by substituting

$$\|\hat{g}_\alpha\|_{L^\infty_{\mathbb{R}}}$$

with

$$\|\hat{g}_\alpha\|_{L^\infty_{[n\Delta\omega, \infty)}}.$$

REMARK 3.9 Observe that the bound of both the quadrature and the cutoff error is given by a product of one factor that depends exclusively on the payoff and another factor that depends on the asset dynamic. This property makes it easy to evaluate the bound for a specific option under different dynamics of the asset price. In Section 4.4, we analyze the terms that depend on the payoff function for the particular case of call options.

REMARK 3.10 From (3.29), it is evident that the speed of the exponential convergence of the trapezoidal rule for analytic functions is dictated by the width of the strip in which the function being transformed is analytic. Thus, in the limit of small error tolerances, it is desirable to set a as large as possible to obtain optimal rates. However, non-asymptotic error tolerances are often practically relevant, and in these cases the trade-off between optimal rates and the constant term $|\hat{g}_\alpha|$ becomes nontrivial. As an example, for the particular case of the Merton model, we have that any finite value of a will do. However, this improvement of the rate of spectral convergence is more than compensated for by the divergence in the constant term.

The integrals in (3.29) and (3.30) can, in some cases, be computed analytically, or bounded from above by a closed-form expression. Consider, for instance, dissipative models with finite jump intensity. These models are characterized by $\sigma^2 > 0$ and $C(\lambda) = 0$. Thus, the integrals can be expressed in terms of the cumulative normal distribution Φ :

$$\int_{\mathbb{R}} e^{-\tau(\sigma^2\omega^2/2)} d\omega = \sqrt{\frac{2\pi}{\tau\sigma^2}}, \quad (3.31)$$

$$\int_{-\infty}^{\infty} e^{-\tau(\sigma^2\omega^2/2)} d\omega = \sqrt{\frac{2\pi}{\tau\sigma^2}}(1 - \Phi(\zeta\sqrt{\tau\sigma^2})). \quad (3.32)$$

Now we consider the case of pure-jump processes (ie, $\sigma^2 = 0$) that satisfy the condition $C(\lambda) > 0$ for some $\lambda \in (0, 2)$. In this case, the integrals are expressible in terms of the incomplete gamma function γ . First, let us define the auxiliary integral:

$$I(a, b) \equiv e^{-a} + a^{-1/b} \gamma\left(\frac{1}{b}, a\right)$$

for $a, b > 0$. Using this, the integrals become

$$\int_{\mathbb{R}} \exp\left(-\tau \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \mathbf{1}_{|\omega|>1}\right) = 2 \left(1 + I\left(\frac{\tau C(\lambda)}{4}, 2 - \lambda\right)\right), \quad (3.33)$$

$$\int_{\varsigma}^{\infty} \exp\left(-\tau \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \mathbf{1}_{|\omega|>1}\right) = \begin{cases} I\left(\frac{\tau C(\lambda)}{4}, 2 - \lambda\right) + 1 - \varsigma, & \varsigma < 1, \\ \varsigma I\left(\frac{\tau \varsigma^{2-\lambda} C(\lambda)}{4}, 2 - \lambda\right), & \varsigma \geq 1. \end{cases} \quad (3.34)$$

An example of a process for which the previous analysis works is the CGMY model, presented in Carr *et al* (2002, 2003), for the regime $Y > 0$.

Last, when both $C(\lambda)$ and σ^2 are positive, the integrals in (3.29) and (3.30) can be bounded by a simpler expression. Consider the two following auxiliary bounds for the same integral, where $\varsigma \geq 1$:

$$\begin{aligned} \int_{\varsigma}^{\infty} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4} C(\lambda)\right)\right) d\omega \\ \leq e^{-\tau(\sigma^2/2)\varsigma^2} \int_{\varsigma}^{\infty} e^{-\tau(|\omega|^{2-\lambda}/4)C(\lambda)} d\omega \\ = \varsigma e^{-\tau(\sigma^2/2)\varsigma^2} I\left(\frac{\tau \varsigma^{2-\lambda} C(\lambda)}{4}, 2 - \lambda\right), \end{aligned} \quad (3.35)$$

$$\begin{aligned} \int_{\varsigma}^{\infty} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4} C(\lambda)\right)\right) d\omega \\ \leq e^{-\tau(\varsigma^{2-\lambda}/4)C(\lambda)} \int_{\varsigma}^{\infty} e^{-\tau(\sigma^2/2)\omega^2} d\omega \\ = \sqrt{\frac{2\pi}{\tau\sigma^2}} e^{-\tau(\varsigma^{2-\lambda}/4)C(\lambda)} (1 - \Phi(\varsigma \sqrt{\tau\sigma^2})). \end{aligned} \quad (3.36)$$

We have that $b(\varsigma)$, defined as the minimum of the right-hand sides of the two previous equations, is

$$b(\varsigma) = \min \left\{ \varsigma e^{-\tau(\sigma^2/2)\varsigma^2} I\left(\frac{\tau \varsigma^{2-\lambda} C(\lambda)}{4}, 2 - \lambda\right), \sqrt{\frac{2\pi}{\tau\sigma^2}} e^{-\tau(\varsigma^{2-\lambda}/4)C(\lambda)} (1 - \Phi(\varsigma \sqrt{\tau\sigma^2})) \right\},$$

a bound for the integral. Bearing this in mind, we have

$$\int_{\mathbb{R}} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4} C(\lambda) \mathbf{1}_{|\omega|>1}\right)\right) d\omega \leq 2\Phi(\sqrt{\tau\sigma^2}) - 1 + 2b(1) \quad (3.37)$$

and

$$\int_{\varsigma}^{\infty} \exp\left(-\tau\left(\frac{\sigma^2}{2}\omega^2 + \frac{|\omega|^{2-\lambda}}{4}C(\lambda)\mathbf{1}_{|\omega|>1}\right)\right) d\omega \leq b(\varsigma), \quad (3.38)$$

provided that $\varsigma \geq 1$.

4 COMPUTATION AND MINIMIZATION OF THE BOUND

In this section, we present numerical examples on the bound presented in the previous section, using practical models known from the literature. We gauge the tightness of the bound compared with the true error using both dissipative and pure-jump processes. We also demonstrate the feasibility of using the expression of the bound as a tool for choosing numerical parameters for the Fourier inversion.

4.1 Call option in VG model

The VG model provides a test case to evaluate the bound in the pure-jump setting. We note that of the two numerical examples presented, it is the less regular model, in the sense that $\sigma^2 = 0$ and $C(\lambda) = 0$ for $0 < \lambda < 2$, indicating that Theorem 3.5 in particular is not applicable.

The Lévy measure of the VG model is given by

$$\nu_{VG}(dy) = dy \left(\mathbf{1}_{y>0} \frac{Ke^{-\eta+y}}{y} - \mathbf{1}_{y<0} \frac{Ke^{\eta-y}}{y} \right),$$

and the corresponding characteristic function is given by Madan *et al* (1998, Equation (7)):

$$\begin{aligned} \varphi \tau \omega &= \left(1 - i\theta \chi \omega + \frac{\sigma^2 \chi}{2} \right)^{-\tau/\chi}, \\ K &= \chi^{-1}, \\ \eta_- &= \left(\sqrt{\frac{\theta^2 \chi^2}{4} + \frac{\sigma^2 \nu}{2}} - \frac{\theta \chi}{2} \right)^{-1}, \\ \eta_+ &= \left(\sqrt{\frac{\theta^2 \chi^2}{4} + \frac{\sigma^2 \nu}{2}} + \frac{\theta \chi}{2} \right)^{-1}. \end{aligned}$$

By Proposition 3.3, we get that

$$a < \min\{\eta_- - \alpha, \eta_+ + \alpha\}, \quad (4.1)$$

which, combined with the requirement that $g_\alpha \in L^1(\mathbb{R})$ (see Remark 3.6), implies

$$\begin{aligned} a &< \min\{\eta_+ - \alpha, \eta_- + \alpha, \alpha - 1\}, \\ a &< \min\{\eta_+ - \alpha, \eta_- + \alpha, -\alpha\} \end{aligned} \quad (4.2)$$

TABLE 1 The error bound for European call/put options in the VG model for select examples.

		K				
		80	90	100	110	120
$12\tau = 1$, $N = 32$	α	-16.9	-13.8	21.6	29.10	36.3
	a	3.33	6.45	18.1	9.77	3.52
	ω_{\max}	229	229	363	363	424
	$\bar{\varepsilon}$	3.35×10^{-4}	0.00334	0.00562	3.97×10^{-4}	7.33×10^{-6}
	$\bar{\varepsilon}^*$	6×10^{-4}	0.0032	0.0058	6×10^{-4}	1×10^{-4}
$12\tau = 4$, $N = 8$	α	-13.8	-13.8	22.1	23.7	29.10
	a	6.11	6.11	17.9	15.2	8.75
	ω_{\max}	62.4	42.4	84.9	126	126
	$\bar{\varepsilon}$	3.99×10^{-4}	0.00312	0.00398	3.57×10^{-4}	1.33×10^{-5}
	$\bar{\varepsilon}^*$	1.3×10^{-3}	0.0057	0.0055	9×10^{-4}	1×10^{-4}

Reference result $\bar{\varepsilon}^*$ from Lee (2004).

for calls and puts, respectively. We note that an evaluation of the integral in (2.13) is also possible for $\alpha \in (0, 1)$ and $\alpha < 0$. In fact, there is a correspondence between shifts in the integration contour and the put–call parity. Integrals with $\alpha < 0$ give rise to put option prices instead of calls. For an extended discussion of this, we refer the reader to Lee (2004) or Boyarchenko and Levendorskii (2011), in which conformal deformation of the integration contour is exploited in order to achieve improved numerical accuracy.

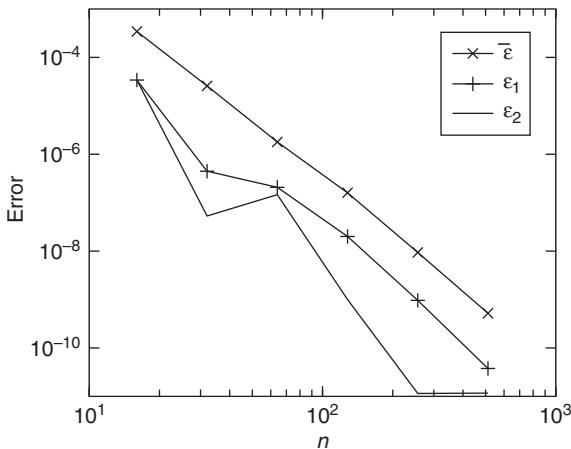
In Lee (2004) and in our calculations, the parameters equal $\eta_+ = 39.7840$, $\eta_- = 20.2648$ and $K = 5.9311$.

Table 1 presents the specific parameters and compares the bound for the VG model with the results obtained by Lee (2004). Based on Table 1, we note that for the VG model presented in Madan *et al* (1998), we can achieve comparable or better error bounds when compared with the study by Lee (2004).

To evaluate the bound, we perform the integration of (3.3) and (3.24) by relying on the Clenshaw–Curtis quadrature method provided in the SciPy package. To supplement Table 1 for a wide range of n , we present the magnitude of the bound compared with the true error in Figure 1.

In Figure 1, we see that the choice of numerical parameters for the Fourier inversion has a strong influence on the error of the numerical method. One does not in general have access to the true solution. Thus, the parameters need to be optimized with respect to the bound. Recall that $\varepsilon = \varepsilon(\alpha, \Delta\omega, a, n)$ and $\bar{\varepsilon} = \bar{\varepsilon}(\alpha, \Delta\omega, n)$ denote the true and estimated errors, respectively. Keeping the number of quadrature points

FIGURE 1 The true error and the error bound for evaluating at-the-money options for the VG model test case.



n fixed, we let $(\alpha_1, \Delta\omega_1, a_1)$ and $(\alpha_2, \Delta\omega_2)$ denote the minimizers of the estimated and true errors, respectively:

$$(\alpha_1, \Delta\omega_1, a_1) = \arg \inf \bar{\mathcal{E}}, \quad (4.3)$$

$$(\alpha_2, \Delta\omega_2) = \arg \inf \mathcal{E}. \quad (4.4)$$

Further, we let \mathcal{E}_1 and \mathcal{E}_2 denote the true error as a function of the parameters minimizing the estimated error and true error, respectively:

$$\mathcal{E}_1 = \mathcal{E}(\alpha_1, \Delta\omega_1), \quad (4.5)$$

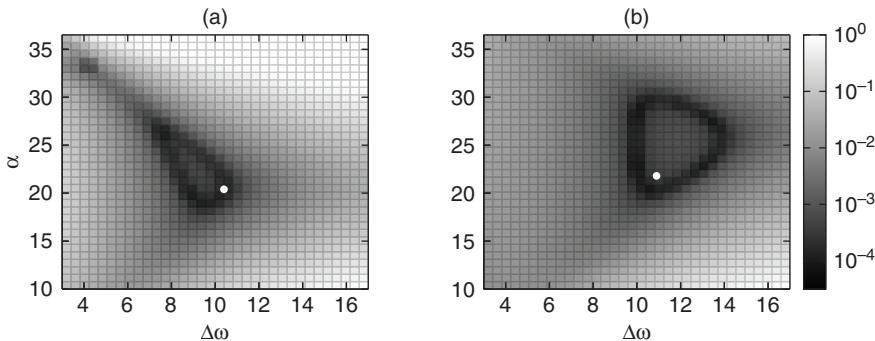
$$\mathcal{E}_2 = \mathcal{E}(\alpha_2, \Delta\omega_2). \quad (4.6)$$

In Figure 1, we see that the true error increases by approximately an order of magnitude when optimizing to the bound instead of to the true error, which translates into a twofold difference in the number of quadrature points needed for a given tolerance. The difference between \mathcal{E}_1 and the bound is approximately another order of magnitude, which necessitates another twofold number of quadrature points compared with the theoretical minimum.

In Figure 2, we present the true error of the Fourier method for the two test cases in Table 1.¹ We note that while minimizing error bounds will produce suboptimal

¹ The reference value for computing the true error was obtained using the numerical methods with n and $\Delta\omega$, such that the level of accuracy is of the order 10^{-10} .

FIGURE 2 The true error ε for the two VG test cases presented in Table 1, and the bound-minimizing configurations (white circle) ($\alpha_2, \Delta\omega_2$) for the examples.



(a) $n = 8$. (b) $n = 32$.

results, the numerical parameters that minimize the bound are a good approximation of the true optimal parameters. This, of course, is a consequence of the error bound having qualitatively similar behavior to the true error, especially as one gets further away from the true optimal parameters.

REMARK 4.1 In practice, the Hardy norm in coefficient M reduces to evaluating an L^1 norm along the two boundaries of the strip of width $2a$. We find that, for practical purposes, the performance of the Clenshaw–Curtis quadrature of the QUADPACK library (provided by the SciPy library) is more than adequate, enabling the evaluation of the bound in a fraction of a second.

For example, the evaluations of the bounds in Table 1 take only around 0.3 seconds on a mid 2014 Macbook Pro equipped with a 2.6 GHz Intel Core i5 processor; this is without attempting to optimize or parallelize the implementation, and while checking for input sanity factors such as the evaluation of the characteristic function in a domain that is a subset in the permitted strip.

We believe that through optimizing routines, skipping sanity checks for inputs and using lower-level computation routines this can be optimized even further, guaranteeing a fast performance even when numerous evaluations are needed.

REMARK 4.2 Like many other authors, we note the exceptional guaranteed accuracy of the FT method with only dozens of quadrature points. This is partially a result of the regularity of the European option price. Numerous Fourier-based methods have been developed for pricing path-dependent options. One might, for the sake of generality, be tempted to use these methods for European options as well, correcting for the

lack of early exercise opportunities. This can be done, certainly; but due to weakened regularity, the required number of quadrature points is easily in the thousands, even when no rigorous bound for the error is required.

We raise one point of comparison, the European option pricing example in Jackson *et al* (2008, Table 2), which indicates a number of quadrature points for pricing the option in the range of thousands. With the method introduced, to guarantee $\mathcal{E} \approx 10^{-3}$, even with no optimization, $n = 64$ turns out to be sufficient.

4.2 Call options under Kou dynamics

To contrast with the pure-jump process presented above, we also test the performance of the bound for the Kou model. We present the relevant results in Table 2. This model differs from the first example in terms of its being dissipative as well as in terms of regularity, in the sense that the maximal width of the domain A_a is, in the case at hand, considerably narrower. The Lévy measure in the Kou model is given by

$$\nu_{\text{Kou}}(dy) = \lambda(p e^{-\eta_1 y} \mathbf{1}_{y>0} + q e^{\eta_2 y} \mathbf{1}_{y<0}),$$

with $p + q = 1$. For the characterization given in Toivanen (2007), the values are set as

$$\begin{aligned}\lambda &= 0.1, & r &= 0.05, & \tau &= 0.25, & S_0 &= 100, \\ p &= 0.3445, & \eta_1 &= 3.0465, & \eta_2 &= 3.0775.\end{aligned}$$

From the expression of the characteristic exponent (see Kou and Wang 2004)

$$\Psi(z) = z \left(r - \frac{\sigma^2}{2} - \lambda \xi \right) + \frac{z^2 \sigma^2}{2} + \lambda \left(\frac{p \eta_1}{\eta_1 - z} + \frac{q \eta_2}{\eta_2 + z} - 1 \right),$$

it is straightforward to see that

$$A_a \subset \{z \in \mathbb{C}: \text{Im } z \in (-3.0465, 3.0775)\}.$$

This range is considerably narrower than that considered earlier. When transforming the option prices in strike space, the relevant expressions for option prices and the error bounds contain a factor exponential in k . The practical implication of this is that, for deep out-of-the-money calls, it is often beneficial to exploit the put–call parity and compute deep in-the-money calls. However, in the case at hand, the strip width does not permit such a luxury. As a consequence, the parameters that minimize the bound are near-identical over a wide range of moneyness, suggesting that we use the FFT algorithm to evaluate the option prices at once for a range of strikes.

TABLE 2 Numerical performance of the bound for the Kou model, with the test case in Toivanen (2007) (see also d'Halluin *et al* 2005), with the number of quadrature points set to $n = 32$.

	K				
	80	90	100	110	120
$\bar{\varepsilon}$	2.67×10^{-4}	3.49×10^{-4}	4.43×10^{-4}	5.52×10^{-4}	6.77×10^{-4}
α	-1.57	-1.57	-1.57	-1.57	-1.57
ω_{\max}	22.9	22.8	22.6	22.5	22.4
$\bar{\varepsilon}^*$	0.34	0.26	0.21	0.17	0.13
$\bar{\varepsilon}^\dagger$	6.87×10^{-4}	1.90×10^{-3}	2.82×10^{-3}	2.72×10^{-3}	2.29×10^{-3}

The point of comparison $\bar{\varepsilon}^*$ refers to the corresponding bound computed with the method described in Lee (2004, Chapters 6.1–6.4). In the $\bar{\varepsilon}^\dagger$, the cutoff error has been evaluated using a computationally more intensive Clenshaw–Curtis quadrature instead of an asymptotic argument with an exponentially decaying upper bound for the option price.

4.3 Binary option in the Merton model

For the particular case of the Merton model, the Lévy measure is given by

$$\nu_{\text{Merton}}(dy) = \frac{\lambda}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(y - r_j)^2}{2\sigma_j^2}\right),$$

and the characteristic exponent is correspondingly given by

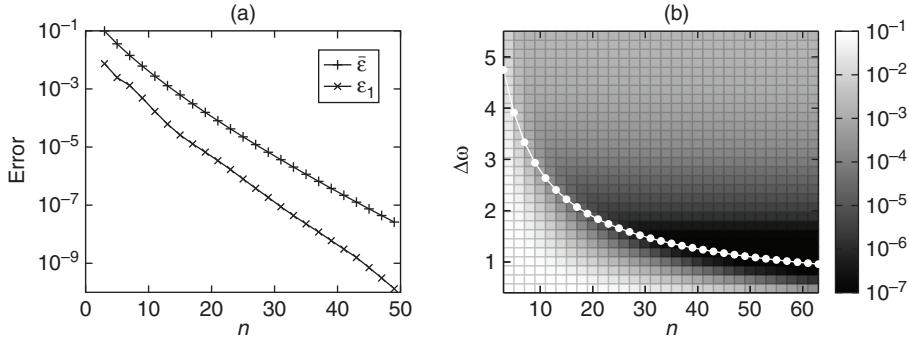
$$\Psi(z)_{\text{Merton}} = z\left(r - \frac{\sigma^2 z}{2}\right) + \frac{\sigma^2 z^2}{2} + \lambda(e^{zr_j + (\sigma_j^2 z^2/2)} - 1 - z(e^{r_j + (\sigma_j^2/2)} - 1)).$$

We may employ a fast, semi-closed-form evaluation of the relevant integrals instead of resorting to quadrature methods. We choose the Merton model as an example of bounding the error of the numerical method for such a model. The parameters are adopted from the estimated parameters for the Standard & Poor's (S&P) 500 index from Andersen and Andreasen (2000):

$$\begin{aligned} S_0 &= 100, & \lambda &= 0.089, & \sigma &= 0.1765, \\ r &= 0.05, & r_j &= -0.8898, & \sigma_j &= 0.4505. \end{aligned}$$

In Figure 3, we present the bound and true error for the Merton model to demonstrate the bound on another dissipative model. The option presented is a binary option with finite support on [95, 105]; no damping was needed or used. We note that, as in the case of the pure-jump module presented above, our bound reproduces the qualitative behavior of the true error. The configuration that results from optimizing the bound is a good approximation of the true error. Such behavior is consistent across the range of n of the most practical relevance.

FIGURE 3 The true error ε_1 and the bound $\bar{\varepsilon}$ for the dissipative Merton model for a range of quadrature points n , along with the bound-minimizing configurations, contrasted with the true error.



4.4 Call options

In Section 3.4, explicit expressions to bound \mathcal{E} are provided. To evaluate these bounds, it is necessary to compute $\|\hat{g}_\alpha\|_{L_\mathbb{R}^\infty}$ and $\|\hat{g}_\alpha\|_{L_{A_\alpha}^\infty}$. According to Remark 3.9, once we compute these values, we could use them for any model, provided that they satisfy the conditions considered there.

The payoff of perhaps the most practical relevance is that of a call option. Consider g , defined by

$$g(x) = (S_0 e^x - K)^+ = S_0(e^x - e^K)^+,$$

for which the selection of a damping parameter $\alpha > 0$ is necessary to have the damped payoff in $L^1(\mathbb{R})$ as well as to ensure the existence of a Fourier transformation. In this case, we have

$$\hat{g}_\alpha(\omega) = S_0 \int_{\mathbb{R}} \exp((1 - \alpha + i\omega)x) - \exp(k + (i\omega - \alpha)x) dx \quad (4.7)$$

$$= \frac{S_0 \exp((1 - \alpha + i\omega)k)}{(1 + i\omega - \alpha)(i\omega - \alpha)} \quad (4.8)$$

and

$$|\hat{g}_\alpha(\omega)|^2 = \frac{S_0^2 e^{2(1-\alpha)k}}{(\alpha^2 + \omega^2)((1 - \alpha)^2 + \omega^2)}. \quad (4.9)$$

It is easy to see that the previous expression decreases as $|\omega|$ increases. This yields

$$\|\hat{g}_\alpha\|_{L_\mathbb{R}^\infty} = |\hat{g}_\alpha(0)| = \frac{S_0 e^{(1-\alpha)k}}{\alpha^2 - \alpha} \quad (4.10)$$

and

$$\|\hat{g}_\alpha\|_{L_{[\zeta, \infty)}^\infty} = |\hat{g}_\alpha(\zeta)|. \quad (4.11)$$

The maximization of $|\hat{g}_\alpha|$ in the strip A_a of the complex plane is more subtle. Denoting $\hat{g}_\alpha(\eta, \rho) = \hat{g}_\alpha(\eta + i\rho)$, we look for critical points that satisfy $\partial_\eta |\hat{g}_\alpha| = 0$. This gives

$$4\eta^3 + 2\eta(4\rho\alpha + 2\alpha^2 - 2\rho - 2\alpha + \rho^2 + 1) = 0. \quad (4.12)$$

For fixed ρ , $|\hat{g}_\alpha|$ has a vanishing derivative with respect to η at a maximum of three points. Of the three roots of the derivative, only the one characterized by $\eta = 0$ is a local maximum; this gives us that, for call options,

$$\|\hat{g}_\alpha\|_{L_{A_a}^\infty} = \sup_{\rho \in [-a, a]} |\hat{g}_\alpha(0, \rho)|. \quad (4.13)$$

Now, observe that $|\hat{g}_\alpha(0, \rho)|$ is a differentiable real function of ρ , whose derivative is given by the following polynomial of second degree:

$$p(\rho) \equiv k(\rho + \alpha - 2\rho\alpha - \alpha^2 - \rho^2) - 2\alpha - 2\rho + 1. \quad (4.14)$$

We conclude that

$$\|\hat{g}_\alpha\|_{L_{A_a}^\infty} = \max_{\rho \in B} \{|\hat{g}_\alpha(0, \rho)|\}, \quad (4.15)$$

where B is the set of no more than four elements consisting of a , $-a$ and the real roots of p that fall in $(-a, a)$.

REMARK 4.3 So far, we have assumed the number of quadrature points n to be constant. In real-life applications, however, this is often not the case. Typically, the user will choose a minimal n that is sufficient to guarantee an error that lies within a predefined error tolerance.

In such a case, we propose the following very simplistic scheme for optimizing numerical parameters and choosing the appropriate n to satisfy an error smaller than ϵ .

- (1) Select $n = n_0$ and optimize to find the relevant configuration.
- (2) See if $\mathcal{E}_Q + \mathcal{E}_F < \epsilon$; if not, increase n by choosing it from a predetermined increasing sequence $n = n_j$, and repeat the procedure.

Especially when using FFT algorithms to evaluate Fourier transforms, we propose $n_j = 2^j n_0$. We further note that, typically, the optimal configuration for the optimizing configuration for n_{j+1} quadrature points does not differ too dramatically from the configuration that optimizes bounds for n_j .

5 CONCLUSION

We have presented a decomposition of the error committed in the numerical evaluation of the inverse Fourier transform, which is needed in asset pricing for exponential Lévy models, into truncation and quadrature errors. For a wide class of exponential Lévy models, we have presented an L^∞ bound for the error.

This error bound differs from the earlier work of Lee (2004) in the sense that it does not rely on the asymptotic behavior of the option payoff at extreme strikes or option prices. This enables the pricing of a wide variety of nonstandard payoff functions, such as those in Suh and Zapatero (2008). The bound, however, does not take into account path-dependent options. We argue that the methods which allow the evaluation of American, Bermudan or knockoff options are considerably more cumbersome and produce significantly larger errors; so, in implementations where performance is important, such as calibration, using American option pricing methods for European options is not justified.

The bound also provides a general framework in which the truncation error is evaluated using a quadrature method; this remains invariant, regardless of the asymptotic behavior of the option price function. The structure of the bound allows for a modular implementation that decomposes the error components arising from the dynamics of the system and the payoff into a product form for a large class of models, including all dissipative models. In select examples, we also demonstrate performances that are comparable or superior to the relevant points of comparison.

We have focused on the minimization of the bound as a proxy for minimizing numerical error. By doing this, one obtains, for a given parameterization of a model, a rigorous L^∞ bound for the error committed in solving the European option price. We have shown that the bound reproduces the qualitative behavior of the actual error. This supports the argument for selecting numerical parameters in a way that minimizes the bound, providing evidence that this selection will, besides guaranteeing numerical precision, be close to the actual minimizing configuration that is not often achievable at an acceptable computational cost.

The bound can be used in the primitive setting of establishing a strict error bound for the numerical estimation of option prices for a given set of physical and numerical parameters, or as a part of a numerical scheme, whereby the end user wishes to estimate an option price either on a single point or in a domain up to a predetermined error tolerance.

In future, the error bounds presented could be used in efforts requiring multiple evaluations of Fourier transformations. Examples of such applications include multi-dimensional Fourier transformations, possibly in sparse tensor grids, as well as time-stepping algorithms for American and Bermudan options. Such applications are

sensitive toward the error bound being used, as any numerical scheme will need to be run multiple times, either in high dimension or for multiple time steps (or both).

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

Häppölä and Crocce are members and Tempone is the director of the KAUST Strategic Research Initiative, Uncertainty Quantification Center. We wish to thank an anonymous referee for their critique, which significantly improved the quality of the manuscript.

REFERENCES

- Almendral, A., and Oosterlee, C. W. (2005). Numerical valuation of options with jumps in the underlying. *Applied Numerical Mathematics* **53**(1), 1–18 (<http://doi.org/fw2zjg>).
- Andersen, L., and Andreasen, J. (2000). Jump-diffusion processes: volatility smile fitting and numerical methods for option pricing. *Review of Derivatives Research* **4**(3), 231–262 (<http://doi.org/bbcqjx2>).
- Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics, Volume 93. Cambridge University Press.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* **24**(1), 1–13 (<http://doi.org/fjwqd4>).
- Black, F., and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3), 637–654.
- Boyarchenko, S., and Levendorskii, S. (2002). Barrier options and touch-and-out options under regular Lévy processes of exponential type. *Annals of Applied Probability* **12**(4), 1261–1298.
- Boyarchenko, S., and Levendorskii, S. (2011). New efficient versions of Fourier transform method in applications to option pricing. Working Paper 1846633, Social Science Research Network.
- Briani, M., Chioma, C. L., and Natalini, R. (2004). Convergence of numerical schemes for viscosity solutions to integro-differential degenerate parabolic problems arising in financial theory. *Numerische Mathematik* **98**(4), 607–646.
- Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: an empirical investigation. *Journal of Business* **75**(2), 305–333 (<http://doi.org/fqvqsv>).
- Carr, P., Geman, H., Madan, D. B., and Yor, M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance* **13**(3), 345–382 (<http://doi.org/d8pqm7>).
- Cont, R., and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.

- De Innocentis, M., and Levendorskii, S. (2014). Pricing discrete barrier options and credit default swaps under Lévy processes. *Quantitative Finance* **14**(8), 1337–1365 (<http://doi.org/btnk>).
- d'Halluin, Y., Forsyth, P. A., and Labahn, G. (2004). A penalty method for American options with jump diffusion processes. *Numerische Mathematik* **97**(2), 321–352 (<http://doi.org/d5sn5v>).
- d'Halluin, Y., Forsyth, P. A., and Vetzal, K. R. (2005). Robust numerical methods for contingent claims under jump diffusion processes. *IMA Journal of Numerical Analysis* **25**(1), 87–112 (<http://doi.org/dkpm7g>).
- Dotsis, G., Psychoyios, D., and Skiadopoulos, G. (2007). An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking and Finance* **31**(12), 3584–3603 (<http://doi.org/d288t3>).
- Dupire, B. (1997). Pricing and hedging with smiles. In *Mathematics of Derivative Securities*, Dempster, S. R., and Pliska, M. A. H. (eds). Cambridge University Press.
- Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. In *Lévy Processes*, pp. 319–336. Springer (<http://doi.org/dgp3sh>).
- Feng, L., and Linetsky, V. (2008). Pricing discretely monitored barrier options and defaultable bonds in Levy process models: a fast Hilbert transform approach. *Mathematical Finance* **18**(3), 337–384 (<http://doi.org/fwbmkb>).
- Hurd, T. R., and Zhou, Z. (2010). A Fourier transform method for spread option pricing. *SIAM Journal on Financial Mathematics* **1**(1), 142–157 (<http://doi.org/dqvtxv>).
- Jackson, K. R., Jaimungal, S., and Surkov, V. (2008). Fourier space time-stepping for option pricing with Lévy models. *The Journal of Computational Finance* **12**(2), 1–29 (<http://doi.org/btnm>).
- Kiessling, J., and Tempone, R. (2011). Diffusion approximation of Lévy processes with a view towards finance. *Monte Carlo Methods and Applications* **17**(1), 11–45 (<http://doi.org/fbx9z>).
- Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science* **48**(8), 1086–1101 (<http://doi.org/fg58sf>).
- Kou, S. G., and Wang, H. (2004). Option pricing under a double exponential jump diffusion model. *Management Science* **50**(9), 1178–1192 (<http://doi.org/bqk52m>).
- Lee, R. W. (2004). Option pricing by transform methods: extensions, unification and error control. *The Journal of Computational Finance* **7**(3), 51–86 (<http://doi.org/btnn>).
- Lord, R., Fang, F., Bervoets, F., and Oosterlee, C. W. (2008). A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes. *SIAM Journal on Scientific Computing* **30**(4), 1678–1705 (<http://doi.org/bd4r4h>).
- Madan, D. B., and Seneta, E. (1990). The variance gamma (VG) model for share market returns. *Journal of Business* **63**(4), 511–524.
- Madan, D. B., Carr, P. P., and Chang, E. C. (1998). The variance gamma process and option pricing. *European Finance Review* **2**(1), 79–105 (<http://doi.org/d4hkvp>).
- Mattner, L. (2001). Complex differentiation under the integral. *Nieuw Archief voor Wiskunde* **2**, 32–35.
- Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* **3**(12), 125–144 (<http://doi.org/d659w5>).

- Nualart, D., and Schoutens, W. (2001). Backward stochastic differential equations and Feynman–Kac formula for Lévy processes, with applications in finance. *Bernoulli* **7**(5), 761–776 (<http://doi.org/cs6qm4>).
- Raible, S. (2000). Lévy processes in finance: theory, numerics, and empirical facts. PhD Thesis, Universität Freiburg im Breisgau.
- Reed, M., and Simon, B. (1975). *Methods of Modern Mathematical Physics, Volume 2: Fourier Analysis, Self-Adjointness*. Academic Press.
- Rydberg, T. H. (1997). The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics: Stochastic Models* **13**(4), 887–910 (<http://doi.org/fmsknc>).
- Schmelzle, M. (2010). Option pricing formulae using Fourier transform: theory and application. Unpublished manuscript.
- Stenger, F. (1993). *Numerical Methods Based on Sinc and Analytic Functions*. Springer.
- Suh, S., and Zapatero, F. (2008). A class of quadratic options for exchange rate stabilization. *Journal of Economic Dynamics and Control* **32**(11), 3478–3501 (<http://doi.org/bd5qf3>).
- Tankov, P. (2004). *Financial Modelling with Jump Processes*. CRC Press.
- Toivanen, J. (2007). Numerical valuation of options under Kou's model. *PAMM* **7**(1), 1024001–1024002 (<http://doi.org/cdssm5>).

Research Paper

Efficient estimation of sensitivities for counterparty credit risk with the finite difference Monte Carlo method

Cornelis S. L. de Graaf,¹ Drona Kandhai^{1,2} and Peter M. A. Sloot^{1,3,4}

¹Computational Science, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands; emails: C.S.L.deGraaf@uva.nl, B.D.Kandhai@uva.nl, P.M.A.Sloot@uva.nl

²Quantitative Analytics, ING Bank, PO Box 1800, 1000 BV Amsterdam, The Netherlands

³National Research University ITMO, 49 Kronverksky Prospekt, St. Petersburg, 197101, Russia

⁴Complexity Institute, S3 B2-A28 Nanyang Avenue, 639798 Singapore

(Received December 16, 2014; revised July 14, 2015; accepted August 24, 2015)

ABSTRACT

According to Basel III, financial institutions have to charge a credit valuation adjustment (CVA) to account for a possible counterparty default. Calculating this measure and its sensitivities is one of the biggest challenges in risk management. Here, we introduce an efficient method for the estimation of CVA and its sensitivities for a portfolio of financial derivatives. We use the finite difference Monte Carlo (FDMC) method to measure exposure profiles and consider the computationally challenging case of foreign exchange barrier options in the context of the Black–Scholes as well as the Heston stochastic volatility model, with and without stochastic domestic interest rate, for a wide range of parameters. In the case of a fixed domestic interest rate, our results show that FDMC is an accurate method compared with the semi-analytic COS method and, advantageously, can compute multiple options on one grid. In the more general case of a stochastic domestic interest rate, we show that we can accurately compute exposures of discontinuous one-touch options by using a linear interpolation

technique as well as sensitivities with respect to initial interest rate and variance. This paves the way for real portfolio level risk analysis.

Keywords: finite difference Monte Carlo (FDMC); credit valuation adjustment (CVA); barrier options; portfolio; exposure computation.

1 INTRODUCTION

Financial crises typically have various causes, but they often have one effect: the call to model more risk factors. Since 1987, we have known that the volatility used in option pricing is not constant and can be better modeled as a stochastic process itself (Heston 1993). More recently, since the Lehman collapse in 2008, measures have been taken to prevent loss of money from a worthless derivative due to counterparty default. Currently, having a single price for an option or financial derivative is therefore not sufficient; institutions also need to know the creditworthiness of their counterparty.

Regulators drafted the Basel III accords (Basel Committee on Banking Supervision 2010), which state that banks need to charge a premium to their trading counterparty for its creditworthiness. This is done via the so-called credit valuation adjustment (CVA), which adjusts the price of a derivative according to the creditworthiness of the counterparty. Moreover, additional capital requirements and limit-monitoring based on potential future losses should be in place. Computing these measures implies that the valuation and risk management of even straightforward plain vanilla options is a high-dimensional and complex problem.

In de Graaf *et al* (2014), we introduced the finite difference Monte Carlo (FDMC) method to calculate the exposure profiles of a derivative. This is done for Bermudan put options, which have an early exercise feature at preset discrete time points. Similar to the methods that appear in Ng and Peterson (2009) and Ng *et al* (2010), the FDMC method uses scenario generation from the Monte Carlo method. Option prices are computed on a grid, and the finite difference method and option values per path are obtained by interpolation on this grid. The expected exposure (EE) equals the mean of the resulting option price distribution, whereas the potential future exposure (PFE) is a quantile of this distribution. In practice, apart from EE and PFE, the sensitivities to market factors (such as spot value, interest rate and volatility) are required for hedging and controlling the counterparty credit risk (CCR) of derivatives portfolios.

In this paper, we extend our previous study. We again incorporate the highly relevant skew effect that is dominantly present in the foreign exchange (FX) market by choosing the Heston model to drive the underlying FX rate. In the case of constant interest rates, we consider the estimation of first- and second-order sensitivities with respect to the spot FX rate. In contrast with the widely used bump-and-revalue method, we propose a path-dependent estimator that is leveraging from the already estimated

local sensitivities on the finite difference grid. A rigorous analysis is performed in the case of barrier options, which pose severe numerical challenges due to the knock-out feature that results in a discontinuous terminal condition. Similar discontinuities also arise in portfolios with instruments of different maturities, with the possibility of error propagation on the computation grid in time. Therefore, we analyze such portfolios specifically in this work. We validate our results by comparing them with the Monte Carlo–COS method (Shen *et al* 2013).

Next, we relax the assumption of a constant domestic interest rate by looking at the three-factor Heston Hull–White model. This implies the use of coarse grids for which the interpolation is vital. We therefore compare exposure quantities for discontinuous one-touch (OT) options computed by a linear or spline interpolation. Additionally, we discuss the applicability of the path-dependent sensitivities with respect to initial variance and domestic interest rate. Again, the bump-and-revalue method acts as a benchmark.

The outline of this paper is as follows. In Section 2, we describe CVA and its sensitivities with respect to the initial underlying values. Section 3 will be the core of this research, where we describe how to use the FDMC method, together with our adjustments, to measure the sensitivity of CVA, and how to extend it to handle multiple options. In Section 4, we present results for a number of test problems; in Section 5, the conclusions are summarized.

2 PROBLEM FORMULATION

2.1 CVA under the Heston Hull–White model

In the Heston Hull–White model, the volatility and the domestic interest rate are modeled as a stochastic process, such that the volatility smile and interest rate dynamics can be captured. The three-dimensional dynamics are given by

$$\begin{aligned} dS_t &= (R_t^d - r^f) S_t dt + \sqrt{V_t} S_t dW_t^1, \\ dV_t &= \kappa(\eta - V_t) dt + \sigma \sqrt{V_t} dW_t^2, \end{aligned} \quad (2.1)$$

$$dR_t^d = \lambda(\theta(t) - R_t^d) dt + \gamma dW_t^3, \quad (2.2)$$

$$dW_t^i dW_t^j = \rho_{i,j} dt \quad \text{for } i \neq j \in [1, 2, 3], \quad (2.3)$$

where r^f is the foreign interest rate, κ is the mean-reverting speed in the Cox–Ingersoll–Ross (CIR) process for the variance, η is the level of the long-term mean and σ is the so-called volatility of volatility. The domestic interest rate R_t^d follows the Hull–White stochastic differential equation (SDE), where λ is the mean-reverting speed, $\theta(t)$ is the level of the long-term mean deduced from the forward curve, and γ is the volatility of the short rate. The SDEs are coupled by the correlated Wiener processes. Note that the same dynamics will hold for an equity derivative with stochastic

interest rate and constant dividend. The FX rate is modeled in more detail if it is also assumed that the foreign interest rate is stochastic, which we do in a forthcoming research paper. The price U of an option with maturity T and payoff function $\phi(S_T, V_T, R_T^d)$, and with the initial value of the underlying volatility and domestic short rate equal to s , v and r^d , respectively, equals

$$U(s, v, r^d, t_0) = \mathbb{E}[e^{-\int_{t_0}^T R_\xi^d d\xi} \phi(S_T, V_T, R_T^d) \mid S_{t_0} = s, V_{t_0} = v, R_{t_0}^d = r^d]. \quad (2.4)$$

Because of the stochastic volatility and interest rate components, pricing formulas are three dimensional, and an analytic option price is harder to obtain, or not available. This is why numerical techniques such as the Monte Carlo or finite difference method are employed to solve the associated partial differential equation (PDE).

For risk purposes, it is obvious that one may be interested in the case where a loss is positive (a negative loss may be a profit); therefore, the exposure of an option at a future time $t < T$ is defined as

$$E(t) := \max(U(S_t, V_t, R_t^d, t), 0), \quad (2.5)$$

where $U(S_t, V_t, R_t^d, t)$ is the value of a financial derivatives contract at time t .

The present EE at a future time $t < T$ is given by

$$\text{EE}(t) := \mathbb{E}[E(t) \mid \mathcal{F}_0], \quad (2.6)$$

where \mathcal{F}_0 is the filtration at time $t = 0$. The discounted version of EE is computed as

$$\text{EE}^*(t) := \mathbb{E}[D(0, t) E(t) \mid \mathcal{F}_0], \quad (2.7)$$

where $D(0, t)$ is the discount factor. In this research, the expectation is calculated under the risk-neutral measure \mathbb{Q} .¹ In the case of a long position in an option, the price (2.4) is always positive; thus, the EE (2.6) is equal to the future option price.

Another important risk assessment is given by the PFE. The quantiles $q = 97.5\%$ and $q = 2.5\%$ of the exposure distribution at time t are defined as

$$\text{PFE}_q(t) = \inf\{x : \mathbb{P}(\text{EE}(t) \leq x) \geq q\}. \quad (2.8)$$

While computing CVA, we assume that the exposure and counterparty's default probability are independent. In the case of the discount factor, exposure and default probability all being independent, we can formulate the expression for CVA as follows (Gregory 2010):

$$\text{CVA}(t_0, T) = (1 - \delta) \int_{t_0}^T \text{EE}^*(t) d \text{PD}(t), \quad (2.9)$$

¹ Typically, the future states can also be modeled under a real-world measure. This is possible when the FDMC method is used, but as this research focuses on the numerical applicability of this method, the risk-neutral measure \mathbb{Q} is assumed.

where δ is the recovery rate and $\text{PD}(t)$ denotes the default probability of the counterparty at time t . The three essential elements are thus recovery rate, discounted EE and default probability.

In practice, CVA is hedged, and thus practitioners compute the sensitivity of the CVA with respect to its dependencies. We assume that the default probability is independent of exposure, such that the sensitivity with respect to Θ (where Θ can be S_0 , V_0 or R_0^d) can be rewritten as

$$\begin{aligned}\frac{\partial \text{CVA}(t_0, T)}{\partial \Theta} &= \frac{\partial}{\partial \Theta} \left((1 - \delta) \int_{t_0}^T \text{EE}^*(t) d \text{PD}(t) \right) \\ &= (1 - \delta) \int_{t_0}^T \frac{\partial \text{EE}^*(t)}{\partial \Theta} d \text{PD}(t).\end{aligned}\quad (2.10)$$

Following the same arguments, the second derivative with respect to Θ can be computed as

$$\begin{aligned}\frac{\partial^2 \text{CVA}(t_0, T)}{\partial \Theta^2} &= \frac{\partial}{\partial \Theta} \left((1 - \delta) \int_{t_0}^T \frac{\partial \text{EE}^*(t)}{\partial \Theta} d \text{PD}(t) \right) \\ &= (1 - \delta) \int_{t_0}^T \frac{\partial^2 \text{EE}^*(t)}{\partial \Theta^2} d \text{PD}(t).\end{aligned}\quad (2.11)$$

By computing these sensitivities in this way, we need an efficient computation of the derivatives $\partial \text{EE}^*(t)/\partial \Theta$ and $\partial^2 \text{EE}^*(t)/\partial \Theta^2$ for every $t \in [t_0, T]$.

To conclude, the CVA of a portfolio is determined by all the future mark-to-market (MtM) values of all the options in the portfolio (Basel Committee on Banking Supervision 2010). Further, if we want to compute the sensitivities, we also need the derivative at all future market scenarios. These requirements call for a valuation method that can compute option prices and derivatives for a wide range of market scenarios. In this paper, we will show that the FDMC method can compute these quantities quickly and accurately.

3 COMPUTATION OF COUNTERPARTY EXPOSURE AND SENSITIVITIES

3.1 The FDMC method

As presented in de Graaf *et al* (2014), the FDMC method uses the scenario generation of the Monte Carlo method and the pricing approach of the finite difference method. The market states are simulated by the quadratic exponential (QE) scheme (Andersen 2008). Next, a grid in the s -, v - and r^d -directions is created. This grid is chosen to be sufficiently large to capture all attained values of the scenario generation. On this grid, prices at any simulation date are calculated by the finite difference procedure.

The specific state (S_m, V_m, R_m^d, t) is interpolated on the grid to obtain the option price $U(S_m, V_m, R_m^d, t)$ at each path, for each time point. At every time point, the resulting future option values for all paths generate a distribution, and from this distribution the exposure profiles can be calculated. The EE can be obtained by averaging over all the prices at all the time points. The higher (97.5%) and lower (2.5%) PFEs can be computed by taking quantiles.

In the case of a path-dependent barrier option, if the underlying state hits the barrier level B , the option is exercised at this path, and the exposure for later time points is set to zero. The essential technique of modeling the exposure by the FDMC method can be presented as follows:

- generate scenarios/paths by Monte Carlo simulation;
- calculate option values and, for barrier options, check which paths hit the barrier;
- set the exposure at each path equal to the option value if the option is not exercised; otherwise, the exposure and all future exposures of this path are set equal to 0;
- compute the empirical distribution of the exposure at each exercise time;
- calculate EE, $\text{PFE}_{2.5\%}$ and $\text{PFE}_{97.5\%}$.

One important difference between the FDMC method and other approaches, such as the regression-based stochastic grid bundling method (SGBM) presented in Feng and Oosterlee (2014) or the Monte Carlo–COS method (Shen *et al* 2013), is that the FDMC method is directly applicable to non-affine models (eg, the stochastic alpha beta rho (SABR) or Heston Hull–White models with nonzero correlation between S_t and R_t^d). To compute exposures driven by non-affine models by the COS method or SGBM, an affine approximation is solved (see Feng and Oosterlee 2014; Guo *et al* 2013). This is not necessary when the FDMC method is used.

3.2 The finite difference method

The risk-neutral value U at $t_0 \leq T$ of a European option with maturity T and payoff function ϕ can be expressed using the conditional expectation under the risk-neutral measure \mathbb{Q} , as follows:

$$U(S_{t_0}, V_{t_0}, R_{t_0}^d, t_0) = \mathbb{E}[e^{-\int_{t_0}^T R_u^d du} \phi(S_T)], \quad (3.1)$$

where $\phi(\cdot)$ is the payoff function of the option. The finite difference procedure computes the price backward in time, starting at maturity $t = T$ and continuing to

$t = t_0$. Thus, the pricing function u is defined as a function of $\tau = T - t$, such that $u(S_\tau, V_\tau, R_\tau^d, \tau) = U(S_{T-t}, V_{T-t}, R_{T-t}^d, T-t)$. The Feynman–Kac theorem links the expectation (3.1) to the solution of a PDE by no-arbitrage arguments. This results in the following PDE:

$$\frac{\partial u}{\partial \tau} = \mathbf{A}u, \quad (3.2)$$

where, in the case of an FX option driven by the Heston Hull–White dynamics, the spatial differential operator \mathbf{A} is given by

$$\begin{aligned} \mathbf{A}u = & \frac{1}{2}vs^2 \frac{\partial^2 u}{\partial s^2} + \frac{1}{2}\sigma^2 v \frac{\partial^2 u}{\partial v^2} + \frac{1}{2}\gamma^2 \frac{\partial^2 u}{\partial (r^d)^2} \\ & + (r^d - r^f)s \frac{\partial u}{\partial s} + (\kappa[\eta - v]) \frac{\partial u}{\partial v} + (\lambda[\theta(T-t) - r^d]) \frac{\partial u}{\partial r^d} \\ & + \rho_{1,2}\sigma vs \frac{\partial^2 u}{\partial s \partial v} + \rho_{1,3}\gamma s \sqrt{v} \frac{\partial^2 u}{\partial s \partial r^d} + \rho_{2,3}\gamma \sigma \sqrt{v} \frac{\partial^2 u}{\partial v \partial r^d} - r^d u. \end{aligned} \quad (3.3)$$

Note that by setting γ and λ equal to 0, we end up with the Heston PDE, where r^d is fixed. For a given state $(S_{\tau_0}, V_{\tau_0}, R_{\tau_0}^d)$ at expiry, the payoff is known; in the finite difference method, this is used as an initial condition.

Barrier options

For a down-and-out barrier call or put option on an underlying S_{τ_0} , with strike K and barrier level B , the payoff function is equal to

$$\phi(S_{\tau_0}) = \max(p(S_{\tau_0} - K), 0)\mathbf{1}_{\{S_{\tau_0} > B\}},$$

with

$$p = \begin{cases} 1 & \text{for a call,} \\ -1 & \text{for a put.} \end{cases} \quad (3.4)$$

The payoff function for European options can be obtained from this by setting $B = 0$.

One-touch options

In the case of OT options, the holder receives a predetermined payout H whenever the underlying reaches K , any time before maturity T . Thus, the payoff at expiry is given as

$$\phi(S_{\tau_0}) = H\mathbf{1}_{\{S_{\tau_0} \geq K\}}. \quad (3.5)$$

3.2.1 Space discretization

In the finite difference method, this PDE is solved on a finite set of points by discretizing in the s -, v - and r^d -directions. The domain to be discretized is chosen as $[0, S_{\max}] \times [0, V_{\max}] \times [-R_{\max}, R_{\max}]$, where S_{\max} , V_{\max} and R_{\max} are chosen to be sufficiently large to minimize the effect of the imposed boundary conditions, and such that all simulated market scenarios can be interpolated on the grid.

Let $s_0 < s_1 < \dots < s_{m_1}$, $v_0 < v_1 < \dots < v_{m_2}$ and $r_0 < r_1 < \dots < r_{m_3}$ be the discretization in the s -, v - and r^d -directions, respectively, similarly to in Haentjens and In't Hout (2012). In all dimensions, the grid is chosen to be nonuniform. The s dimension consists of a predefined interval $[S_{\text{left}}, S_{\text{right}}]$, in which points are uniformly spaced. S_{left} and S_{right} are chosen to contain the region of interest, ie, the region around the expected mean of the underlying. Following Haentjens and In't Hout (2012), for options without barriers we choose

$$[S_{\text{left}}, S_{\text{right}}] = [0.5K, K].$$

Outside $[S_{\text{left}}, S_{\text{right}}]$, the points are distributed with the help of a hyperbolic sine function. In the barrier case, the nonuniform grid is chosen such that the dense region contains more than 95% of the non-exercised paths; generally, choosing

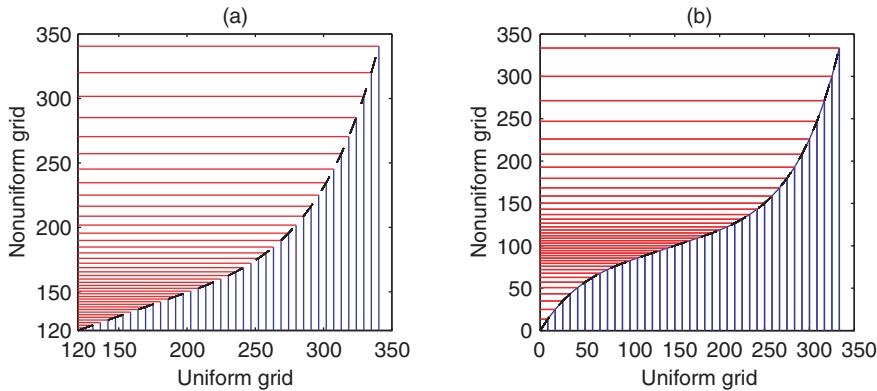
$$[S_{\text{left}}, S_{\text{right}}] = \begin{cases} [0.5K, B] & \text{for a up-and-out call or put,} \\ [B, 1.5K] & \text{for a down-and-out call or put} \end{cases}$$

is found to be sufficient. For a portfolio of options, however, we define S_{left} and S_{right} such that all possible strikes and barriers are included. In Figure 1, the two different nonuniform grids are shown.

In the v -direction, the grid is chosen similarly to in In't Hout and Foulon (2010). The grid is dense around $v = 0$. We do this because, for realistic test parameters, the expected mean of the variance process is close to 0. In addition, because the Heston PDE in the v -direction is convection-dominated close to 0 and the initial condition is non-smooth, numerical stability requires a high density of points in this region (Haentjens and In't Hout 2012).

Also, for the r^d -direction, we follow Haentjens and In't Hout (2012). The grid is dense around 0 and stretched symmetrically toward the boundaries $-R_{\max}$ and R_{\max} by using a sinus hyperbolic function.

The derivatives are approximated using central, forward and backward three-point stencils. All stencils are second-order accurate. For more details, we refer to Haentjens and In't Hout (2012).

FIGURE 1 Nonuniform grids in the s -direction.

(a) A down-and-out put (DOP) option (where we choose $S_{\text{left}} = B = 120$ and $S_{\text{right}} = 140$). (b) A portfolio of options (with discontinuous points within $S_{\text{left}} = 100$ and $S_{\text{right}} = 150$).

TABLE 1 Boundary conditions and payoff functions under the Black–Scholes dynamics.

Option type	$s \rightarrow S_{\max}$	$s \rightarrow S_{\min}$
European call	$\frac{\partial u}{\partial s} = 1$	$u = 0$
European put	$u = 0$	$u = e^{-\int_{\tau_0}^{\tau} r_u^d du} K_P$
Up-and-out barrier call	$u = 0$	$u = 0$
Up-and-out barrier put	$u = 0$	$u = e^{-\int_{\tau_0}^{\tau} r_u^d du} K_P$
Down-and-out barrier call	$\frac{\partial u}{\partial s} = 1$	$u = 0$
Down-and-out barrier put	$u = 0$	$u = 0$
One-touch	$u = H$	$u = 0$

3.2.2 Boundary conditions

The options considered in this research are of the following type: European call and put options with strike K_C and K_P , respectively; barrier options with strike K_B and barrier level B , which can be down-and-out or up-and-out calls or puts; and OT options with strike K_O and payout H . The boundary conditions for the s dimension used in this research are stated in Table 1. Note that in the case of non-barrier options, S_{\min} and S_{\max} converge to 0 or ∞ , respectively.

The boundary conditions in the volatility direction are imposed independently of the option type. In Ekström and Tysk (2011), it is shown that for a CIR process,

such as the variance process in the Heston model, the solution of the PDE at $v = 0$ satisfies the boundary condition that is obtained by inserting $v = 0$ into (3.3); this is also referred to as a degenerated boundary condition:

$$\frac{\partial u}{\partial \tau} = \kappa \eta \frac{\partial u}{\partial v} + (r^d - r^f) s \frac{\partial u}{\partial s} - r^d u. \quad (3.6)$$

In traditional literature (see, for example, Tavella and Randall 2000), the maximum variance boundary for call options is imposed as $u(s, V_{\max}, r^d, \tau) = s$, but experiments show that this introduces a boundary layer. In combination with the PDE becoming convection-dominated around $v \approx 0$, this can result in oscillations if no upwinding is applied. To prevent this problem but still use central schemes, the option value at the maximum variance boundary is assumed to satisfy

$$\frac{\partial^2 u(s, V_{\max}, r^d, \tau)}{\partial v^2} = 0. \quad (3.7)$$

For the interest rate dimension, which is needed in the Heston Hull–White model, the boundary conditions are taken as

$$\frac{\partial u(s, v, \pm R_{\max}^d, \tau)}{\partial r^d} = 0. \quad (3.8)$$

Using the discretizations, boundary and initial condition, the following initial value problem for stiff ordinary differential equations (ODEs) is derived:

$$\left. \begin{aligned} \mathbf{u}'(\tau) &= A\mathbf{u}(\tau) + \mathbf{g}(\tau), \\ \mathbf{u}(\tau_0) &= \phi(s(T)), \end{aligned} \right\} \quad (3.9)$$

where $\mathbf{u}(\tau)$ denotes the vector of discrete solutions $\mathbf{u}_{i,j,k}(\tau) := u(s_i, v_j, r_k^d, \tau)$ ordered lexicographically, $\mathbf{g}(\tau)$ is a vector determined by the boundary conditions and $s(T)$ denotes the grid in the s -direction at maturity.

3.2.3 Time discretization

As the Heston model is a two-dimensional problem in space, the ODEs also have two space dimensions. To solve problems with dimensions higher than one, splitting techniques are relevant.

The splitting scheme used in this research is the Hundsdorfer–Verwer scheme. For more details, we refer to Hundsdorfer and Verwer (2003) for the derivation of the scheme, and to In’t Hout and Welfert (2009) for a more detailed explanation of the alternating direction implicit (ADI) schemes in this context.

3.3 Computing CVA and its sensitivities

To estimate CVA, we need $\text{EE}(t)$ at any time $t \in [t_0, T]$ during the life of the derivative. Next to that, we need the probability of default at any time. Following Gregory (2010), we define $q_l = q(t_{l-1}, t_l)$ as the probability that the counterparty will default in the interval $[t_l - dt, t_l]$. Using the so-called hazard rate λ_{haz} , the survival probability $P_{\text{surv}}(t)$ is defined as

$$P_{\text{surv}}(t) := e^{-\lambda_{\text{haz}} t}. \quad (3.10)$$

Using this definition, we can derive the probability to default in interval $(t - dt, t)$ conditioned on no prior default, as follows:

$$q(t - dt, t) = P_{\text{surv}}(t) - P_{\text{surv}}(t - dt). \quad (3.11)$$

For any counterparty for which a credit default swap (CDS) is available for protection, this entity can be calculated from the CDS spread. As shown in Whetten *et al* (2004), the annual premium payment c of a CDS can be calculated as

$$c = \frac{(1 - \delta) \sum_{l=1}^N P(t_0, t_l)(q_{l-1} - q_l)}{\sum_{l=1}^N P(t_0, t_l)q_l dt + \sum_{l=1}^N P(t_0, t_l)(q_{l-1} - q_l)(dt/2)}, \quad (3.12)$$

where dt denotes the payment interval. In this research, we assume annual premiums of 400 basis points (bps), which correspond to a hazard rate of 6.6×10^{-2} . Now, in a discrete setting, CVA can be calculated as

$$\text{CVA} = (1 - \delta) \sum_{l=1}^N q(t_{l-1}, t_l) \text{EE}^*(t_l). \quad (3.13)$$

By using this expression, the first and second derivative of the CVA with respect to Θ_0 (where Θ_0 can be S_0 , V_0 and R_0^d) can be derived as follows:

$$\begin{aligned} \frac{\partial \text{CVA}}{\partial \Theta_0} &= \frac{\partial}{\partial \Theta_0} (1 - \delta) \sum_{l=1}^N q(t_{l-1}, t_l) \text{EE}^*(t_l), \\ &= (1 - \delta) \sum_{l=1}^N q(t_{l-1}, t_l) \frac{\partial \text{EE}^*(t_l)}{\partial \Theta_0}, \end{aligned} \quad (3.14)$$

where, in the second equality, we assume independence between the default probability and Θ_0 as well as between the recovery rate δ and Θ_0 . Note that the assumption of independence between the default probability and Θ_0 can be relaxed by modeling the default probability as a stochastic process that depends on Θ_0 , as is done in Hull

and White (2012).² Similarly, for the second derivative, we have

$$\frac{\partial^2 \text{CVA}}{\partial \Theta_0^2} = (1 - \delta) \sum_{l=1}^N q(t_{l-1}, t_l) \frac{\partial^2 \text{EE}^*(t_l)}{\partial \Theta_0^2}. \quad (3.15)$$

To compute

$$\frac{\partial \text{EE}^*(t)}{\partial \Theta_0}$$

in (3.14), first, the derivative is rewritten as follows:

$$\frac{\partial \text{EE}^*(t)}{\partial \Theta_0} = \frac{\partial \text{EE}^*(t)}{\partial S_t} \frac{\partial S_t}{\partial \Theta_0} + \frac{\partial \text{EE}^*(t)}{\partial V_t} \frac{\partial V_t}{\partial \Theta_0} + \frac{\partial \text{EE}^*(t)}{\partial R_t^d} \frac{\partial R_t^d}{\partial \Theta_0}. \quad (3.16)$$

At every intermediate time point t_l , the finite difference method stores the prices for the entire grid in the vector $\mathbf{u}^l = \text{EE}^*(t_l)$. On this grid, we can approximate $\partial \text{EE}^*(t_l)/\partial \Theta_t$ and $\partial^2 \text{EE}^*(t_l)/\partial \Theta_t^2$ by multiplying with the difference matrixes $\mathbf{A}_{\Theta_t}(t)$ and $\mathbf{A}_{\Theta_t^2}(t)$. These are defined as follows:³

$$\frac{\partial \text{EE}^*(t_l)}{\partial \Theta_t} \approx \mathbf{A}_{\Theta_t}(t_l) \mathbf{u}^l = \frac{\partial \mathbf{u}(t_l)}{\partial \Theta_t} + \mathcal{O}(\Delta \Theta_t^2), \quad (3.17)$$

$$\frac{\partial^2 \text{EE}^*(t_l)}{\partial \Theta_t^2} \approx \mathbf{A}_{\Theta_t^2}(t_l) \mathbf{u}^l = \frac{\partial^2 \mathbf{u}(t_l)}{\partial \Theta_t^2} + \mathcal{O}(\Delta \Theta_t^2). \quad (3.18)$$

So, the partial derivatives of the exposure in (3.16) are obtained from the finite difference grid. The partial derivatives of the state variables with respect to the initial conditions are analyzed in the following subsections for all the possible choices of Θ .

3.3.1 Sensitivity with respect to initial FX rate

To compute the sensitivity with respect to the initial underlying FX rate ($\Theta_0 = S_0$), we first note that the future variance and future short rate are independent of S_0 , such that $\partial V_t/\partial S_0 = \partial R_t^d/\partial S_0 = 0$. However, $\partial S_t/\partial S_0$ is clearly nonzero, and this can be computed by the pathwise Monte Carlo method. Because S_0 follows a geometric Brownian motion (GBM) in the Heston Hull–White model, we can assume

$$S_t = S_0 e^{(R_t^d - r^f - (V_t/2))t + \sqrt{V_t} \sqrt{t} Z}, \quad (3.19)$$

² In this specific case, (3.14) will have an extra term, but the sensitivities can still be computed. We leave this application for future work.

³ Note that for $\Theta_t = r_t^d$, $\mathbf{A}_r(t)$ is a time-dependent matrix; as in the case of the stochastic interest rate, the drift can be time dependent because of the yield curve (see, for example, the Hull–White model).

where Z is a standard normal random variable. Consequently, following Broadie and Glasserman (1996), for the first and second derivative, we have

$$\frac{\partial S_t}{\partial S_0} = e^{(R_t^d - r^f - V_t/2)t + \sqrt{V_t}\sqrt{t}Z} = \frac{S_t}{S_0}, \quad (3.20)$$

$$\frac{\partial^2 S_t}{\partial S_0^2} = 0. \quad (3.21)$$

Now, at any time point t_l , both partial derivatives from (3.16) can be computed for every path, such that $\partial \text{EE}^*(t)/\partial S_0$ is obtained by averaging.

To compute $\partial^2 \text{CVA}/\partial S_0^2$, we need $\partial^2 \text{EE}^*(t)/\partial S_0^2$; this yields

$$\begin{aligned} \frac{\partial^2 \text{EE}^*(t)}{\partial S_0^2} &= \frac{\partial}{\partial S_0} \left(\frac{\partial \text{EE}^*(t)}{\partial S_t} \frac{\partial S_t}{\partial S_0} \right) \\ &= \left(\frac{\partial}{\partial S_0} \frac{\partial \text{EE}^*(t)}{\partial S_t} \right) \frac{\partial S_t}{\partial S_0} + \frac{\partial \text{EE}^*(t)}{\partial S_t} \left(\frac{\partial}{\partial S_0} \frac{\partial S_t}{\partial S_0} \right) \\ &= \left(\frac{\partial^2 \text{EE}^*(t)}{\partial S_t^2} \frac{\partial S_t}{\partial S_0} \right) \frac{\partial S_t}{\partial S_0} + \frac{\partial \text{EE}^*(t)}{\partial S_t} \frac{\partial^2 S_t}{\partial S_0^2} \\ &= \frac{\partial^2 \text{EE}^*(t)}{\partial S_t^2} \left(\frac{S_t}{S_0} \right)^2, \end{aligned} \quad (3.22)$$

where the second derivative of EE with respect to S_t can be obtained from (3.18), and $(S_t/S_0)^2$ can be obtained from the scenario generation.

3.3.2 Sensitivity with respect to initial variance

In the case of sensitivity with respect to initial variance ($\Theta_0 = V_0$), the future short rate is independent of V_0 , such that (3.16) can be simplified to

$$\frac{\partial \text{EE}^*(t)}{\partial V_0} = \frac{\partial \text{EE}^*}{\partial S_t} \frac{\partial S_t}{\partial V_0} + \frac{\partial \text{EE}^*}{\partial V_t} \frac{\partial V_t}{\partial V_0}. \quad (3.23)$$

The partial derivatives of the exposures can be extracted from the finite difference grid by (3.17), such that only $\partial S_t/\partial V_0$ and $\partial V_t/\partial V_0$ are unknown. In the Heston model, the variance is modeled by a square root process, which has its difficulties. In this model, the variance process can and will reach zero, such that a straightforward derivative of (3.19) with respect to V_0 is not defined for every path at every time. This also holds for $\partial V_t/\partial V_0$, because when the discretized SDE of the variance process is differentiated with respect to V_0 , the square root will appear in the denominator, which makes the derivative intractable. Further, in Chan and Joshi (2010) it is noted that the sensitivities of the variance process with respect to initial inputs can grow very quickly and potentially blow up. We therefore approximate these partial derivatives

by a local bump-and-revalue approach, as follows:

$$\frac{\partial S_t}{\partial V_0} \approx \frac{\tilde{S}_m - S_m}{\epsilon_v}, \quad (3.24)$$

$$\frac{\partial V_t}{\partial V_0} \approx \frac{\tilde{V}_m - V_m}{\epsilon_v}, \quad (3.25)$$

where \tilde{S} and \tilde{V} are modeled with $V_0 + \epsilon_v$ as the initial variance. Note that when an analytic expression for these partial derivatives is available, the method will gain in efficiency.

3.3.3 Sensitivity with respect to initial domestic interest rate

To measure the sensitivity with respect to the initial domestic short rate R_0^d , (3.16) is simplified to

$$\frac{\partial \text{EE}^*(t)}{\partial r_0^d} = \frac{\partial \text{EE}^*}{\partial S_t} \frac{\partial S_t}{\partial R_0^d} + \frac{\partial \text{EE}^*}{\partial R_t^d} \frac{\partial R_t^d}{\partial R_0^d}, \quad (3.26)$$

where $\partial V_t / \partial R_0^d$ is 0 because the future variance is independent of the initial short rate. Similar to the previous cases, the partial derivatives of EE with respect to S_t and R_t^d can be derived from the finite difference grid by (3.17).

The partial derivatives of the state variables with respect to R_0^d can be computed along the path in the Monte Carlo simulation. For any discrete time point t_l ($0 < l < N$), we can create a recursive formula for $\partial S_{t_l} / \partial R_0^d$:

$$\frac{\partial S_{t_l}}{\partial R_0^d} = \frac{\partial S_{t_l}}{\partial S_{t_{l-1}}} \frac{\partial S_{t_{l-1}}}{\partial R_0^d} + \frac{\partial S_{t_l}}{\partial R_{t_{l-1}}} \frac{\partial R_{t_{l-1}}^d}{\partial R_0^d}, \quad (3.27)$$

where (when S_t is driven by a GBM) we have

$$\frac{\partial S_{t_l}}{\partial S_{t_{l-1}}} = \frac{S_{t_l}}{S_{t_{l-1}}}, \quad (3.28)$$

$$\frac{\partial S_{t_l}}{\partial R_{t_{l-1}}^d} = S_{t_l} \Delta t. \quad (3.29)$$

Here, $\Delta t = t_l - t_{l-1}$ is the uniform time increment in one time step. The interest rate is modeled by the Hull–White model (Hull and White 1993), such that a Euler scheme as a discretization yields

$$R_{t_{l+1}}^d = R_{t_l}^d + \lambda[\theta(t_l) - R_{t_l}^d]\Delta t + \gamma\sqrt{\Delta t}Z_{t_l}, \quad (3.30)$$

where $Z \sim N(0, 1)$. From this, we can recursively derive

$$\frac{\partial R_{t_{l-1}}}{\partial R_0^d} = (1 - \lambda\Delta t)^{(l-1)}. \quad (3.31)$$

The first time step gives us the initial condition

$$\frac{\partial S_{t_1}}{\partial R_0^d} = S_0 \Delta t. \quad (3.32)$$

Using this recursive formula, together with the finite difference approximations, we can estimate the sensitivity with respect to R_0^d at any time point, without the need of an extra Monte Carlo simulation.

Similar to the case for EE, the computation of the first and second derivatives with respect to initial underlying states can be summarized as follows:

- generating scenarios/paths by Monte Carlo simulation;
- at each time point t_l , for the entire grid, calculate option sensitivities $\partial \text{EE}^*(t_l)/\partial \Theta$ and $\partial^2 \text{EE}^*(t_l)/\partial \Theta^2$; for barrier options, check if the option is not exercised ($S_{t^*} < B$ for all $t^* \leq t_l$);
- set the first and second derivatives at each path as the calculated sensitivities if the option is not exercised; otherwise, set them equal to 0;
- compute the empirical distribution of the sensitivities at each exercise time;
- calculate $\partial \text{EE}^*(t_l)/\partial \Theta$ and $\partial^2 \text{EE}^*(t_l)/\partial \Theta^2$ by averaging.

3.4 Pricing a portfolio

In this research, the finite difference grid is used to price multiple options with different strikes and maturities in one sweep on one grid. The portfolios considered here are constructed of European options and a first-order exotic barrier option. The value Π of a portfolio of N options can be seen as the sum of the option prices

$$\Pi(t) = \sum_{i=1}^N U_i(S_t, K_i, T_i), \quad (3.33)$$

where K_i is the strike, T_i is the maturity and U_i is the price of option i . In this paper, option i can thus be a European call or put option, or a barrier option. We assume that all options in the portfolio can be netted.

Together with the Monte Carlo scenario generation, this gives us the exposure profile of the sum of the option values at any future time point. The duration of the portfolio is equal to the longest maturity in the portfolio:

$$\tilde{T} = \max_{i \in [1, N]} T_i. \quad (3.34)$$

Again, at this maximum maturity, all the option prices on the grid are known; therefore, the time is reversed such that the payoff formula (3.35) can be used as an initial

condition that equals the sum of all individual payoff functions belonging to options with maturity equal to the maximum maturity \tilde{T} :

$$\phi_P(S_t) = \sum_{i=1}^N \phi_i(S_t, K_i, T_i) \mathbf{1}_{\{T_i=\tilde{T}\}}. \quad (3.35)$$

Important in the context of this research is that the option-specific characteristics are only introduced by the initial condition and the boundary conditions. Because the portfolio consists of a sum of options, the boundary conditions for the portfolio will also be just a sum of these limiting conditions, such that our time-stepping routine in the finite difference procedure can be updated as follows.

- From \mathbf{u}^l , calculate $\tilde{\mathbf{u}}^{l-1}$ by the ADI splitting scheme.
- Update the portfolio value with possible other option values:

$$\mathbf{u}^{l-1} = \tilde{\mathbf{u}}^{l-1} + \sum_{i=1}^N \phi_i(s, K_i, T_i) \mathbf{1}_{\{T_i=t_{l-1}\}},$$

where s is a vector of the same size as \mathbf{u} , consisting of all s grid points.

- Update all boundary conditions.
- If $l > 0$, repeat the procedure.

By applying this time-stepping procedure, \mathbf{u}^0 will be the value of the portfolio at time $t = t_0$. For the computation of exposure of a portfolio over time, options that are not path-dependent can be included on one grid. By using only one grid, there will be no extra computational time for these extra options.

In the case of a portfolio of a call, put and barrier considered in this research, the EEs of the call and put options are computed using only one grid, whereas the EE of the barrier option is computed using the algorithm from Section 3.1. In general, for the computation of path-dependent options, first a separate finite difference procedure needs to be done. Next, while computing this individual options exposure, it needs to be checked, for any scenario, if the simulated scenario should be exercised or not. At every time step, the portfolio exposure is then computed as the sum of the individual barrier exposure and the call and put option exposure. To summarize, the computational time of computing exposure for a portfolio is determined by

- one Monte Carlo simulation to compute all the scenarios,
- one finite difference procedure to compute the price grid for all non-path-dependent options,
- a separate finite difference procedure per path-dependent option.

Note that the same holds for path-dependent American options.

TABLE 2 Model parameters for various test cases.

	Case A	Case B
Spot (S_0)	1.364	138.1
Foreign short rate (r^f)	0.01	0.10
Initial variance (V_0)	0.029	0.029
Mean reversion speed of variance (κ)	4.42	1.50
Mean reversion variance level (η)	0.0240	0.0707
Volatility of volatility (σ)	0.46	0.63
Initial domestic short rate (R_0^d)	0.01	0.03
Mean reversion speed of interest rate (λ)	0.5	0.5
Mean reversion interest rate (θ)	0.05	0.05
Volatility of short rate (γ)	0.02	0.02
s, v -correlation ($\rho_{1,2}$)	-0.45	-0.76
s, r^d -correlation ($\rho_{1,3}$)	0.501	-0.011
v, r^d -correlation ($\rho_{2,3}$)	-0.96	-0.96
Maturity (T)	0.5	1.0
Strike (K)	1.360	138.1
Barrier (B)	1.20	120

4 NUMERICAL RESULTS

The numerical results are divided into two parts. First, we present our numerical study on the accuracy and convergence under the two-dimensional Heston model. In this case, all results are validated by the semi-analytical COS method.

Second, we further assess the accuracy of the interpolation schemes by looking into the numerically challenging Heston Hull–White model, where the number of grid points per dimension is smaller, which can cause a larger interpolation error. We do this in combination with the EE computation of OT options, which are discontinuous during the entire lifetime of the option.⁴ Next to that, we show the sensitivities with respect to V_0 and R_0^d .

The parameters are chosen according to Table 2. In case A, the foreign interest rate is equal to the initial domestic rate; next to that, the option is out-of-the-money (OTM) at inception. The level of the initial FX rate is set to 1.3639, which is a real market-quoted EUR/USD FX rate from June 2014, whereas the other parameters satisfy characteristics observed in literature, such as negative correlation, low volatility and interest rates and small maturities (see, for example, Schoutens *et al* 2004 and Albrecher *et al* 2007). In this test, the well-known Feller condition is satisfied. In case B, the initial FX rate is set to 138.1, a real EUR/JPY FX rate from June 2014;

⁴ We would like to thank an anonymous reviewer for this suggestion.

the option is at-the-money (ATM) at inception, while the initial domestic interest rate is higher than the foreign interest rate. In this case, the other model parameters are chosen such that the Feller condition is violated.⁵

4.1 Heston model

In the case of the Heston model, the domestic interest rate R_t^d is assumed to be constant over time, such that

$$D(0, t) = e^{-\int_{t_0}^t R_\xi^d d\xi} = e^{-r^d t}.$$

Further, in the pricing PDE (3.3), λ and γ are assumed to be 0.

4.1.1 Single barrier options: numerical setup

Computing the exposure of barrier options is more challenging than computing the exposure of European options. Barrier options are path-dependent and have a discontinuous initial condition. It is this discontinuous nature of the payoff function in particular that may complicate the accurate estimation of sensitivities, especially for higher-order ones. Because we have a benchmark solution for down-and-out put (DOP) options, we do an extensive error analysis for this option type, but the method can also be applied to down-and-out call (DOC), up-and-out put (UOP) and up-and-out call (UOC) options, and all the other “in” (instead of “out”) variants.

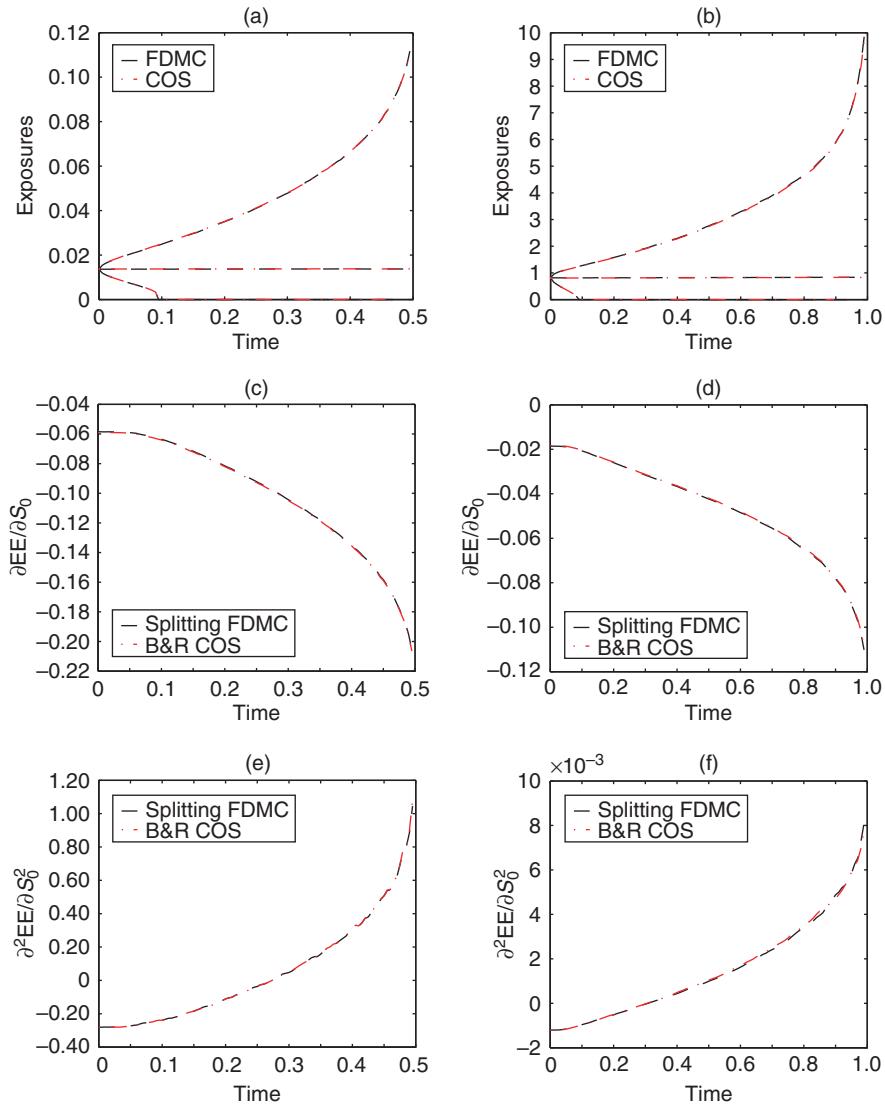
The computed EE, PFE_{2.5%} and PFE_{97.5%} are shown in Figures 2(a) and 2(b). The starting level of the EE equals the option price at $t = t_0$ and shows a small increase toward maturity. The PFE, however, shows more interesting behavior. Starting at the option price, the PFE is increasing over time and shows a steep growth close to maturity. Intuitively, the increase of the higher quantile makes sense; when moving $t^* \in [t_0, T]$ closer to maturity, the hitting probability conditioned on no prior barrier hit will become smaller, such that for in-the-money paths the price will resemble a straightforward European option value more and more. The mean (EE) is not heavily affected, because the probability of the barrier being hit up to time t^* is increased, which will lower the option value.

4.1.2 Accuracy and convergence

As a benchmark, the COS method can be applied to evaluate barrier options accurately and efficiently. For details on the pricing procedure using this Fourier cosine method, we refer to Fang and Oosterlee (2011). Here, we use this efficient pricing technique by computing prices for an entire grid of possible market scenarios. Similar to the grid used for the finite difference procedure, this grid is chosen so that it is large enough to

⁵ When the Feller condition is not satisfied, the variance process can become zero, and numerical methods can become unstable.

FIGURE 2 Exposures (EE, PFE_{2.5%} and PFE_{97.5%}) and the first and second derivative profiles over time under the Heston dynamics for tests A and B.



The dashed black line is computed using the FDMC method, whereas the dashed red line is computed using the COS method. In the case of the sensitivities, the results corresponding to the COS method are obtained using a bump-and-revalue (B&R) procedure, whereas for the FDMC method the derivative is splitted, as explained in Section 3.3. (a) Exposure profiles for test A. (b) Exposure profiles for test B. (c) Delta profile for test A. (d) Delta profile for test B. (e) Gamma profile for test A. (f) Gamma profile for test B.

contain all future market scenarios generated by the Monte Carlo scenario generation. We choose 500 points in the s -direction and 300 points in the v -direction, and both are densely distributed around the expected means. Prices for all the scenarios are obtained by a spline interpolation on the COS grid.

For the sensitivities with respect to S_0 , we run this procedure two (in the case of Delta) or three (in the case of Gamma) times, with an initial condition bumped by ϵ_s . From the resulting EEs, the sensitivities are computed using the following finite difference formulas:

$$\frac{\partial \text{EE}(t)}{\partial S_0} \approx \frac{\text{EE}_{S_0+\epsilon_s}(t) - \text{EE}_{S_0-\epsilon_s}(t)}{2\epsilon_s}, \quad (4.1)$$

$$\frac{\partial^2 \text{EE}(t)}{\partial S_0^2} \approx \frac{\text{EE}_{S_0+\epsilon_s}(t) - 2\text{EE}_{S_0}(t) + \text{EE}_{S_0-\epsilon_s}(t)}{\epsilon_s^2}. \quad (4.2)$$

Figures 2(a) to 2(f) show that the exposure profiles and sensitivities over time computed with the FDMC method resemble the results computed by the Monte Carlo–COS method.

For an EE computed over N_T evaluation dates, the relative L_2 and L_∞ errors are computed as

$$\|\cdot\|_\infty := \frac{\max_{i=1,\dots,N_T} |\text{EE}_i^{\text{COS}} - \text{EE}_i^{\text{FDMC}}|}{\max_{i=1,\dots,N_T} |\text{EE}_i^{\text{COS}}|}, \quad (4.3)$$

$$\|\cdot\|_2 := \frac{(\sum_{i=1}^{N_T} (\text{EE}_i^{\text{COS}} - \text{EE}_i^{\text{FDMC}})^2)^{1/2}}{(\sum_{i=1}^{N_T} (\text{EE}_i^{\text{COS}})^2)^{1/2}}. \quad (4.4)$$

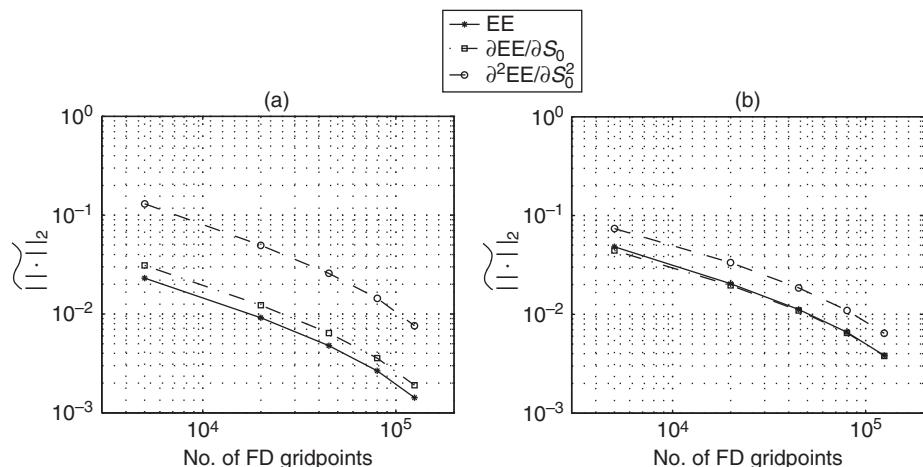
In Table 3, the errors between the FDMC method with 700 grid points in the s -direction and 350 in the v -direction are compared with the COS method. We can see that the relative error is below 1% in both EE and the first derivative. The second derivative, however, is accurate up to 5% in both the L_∞ and L_2 norms. This is due to the fact that this absolute value of Gamma is already in the range of 10^{-4} , such that the errors from the finite difference discretization have a larger impact. Further, we can see that the difference between a spline and linear interpolation is negligible.

The convergence with respect to the number of finite difference grid points is shown in Figures 3(a) and 3(b) for the EE and the first and second derivative with respect to S_0 . In this case, the benchmark is the converged finite difference solution obtained with 700 and 350 points in the s - and v -directions, respectively. The convergence is shown to be first-order in the number of grid points.

In Figures 4(a) and 4(b), we show the decline of the relative standard error (SE) in percentage of the mean by increasing the number of paths for tests A and B. Here, we computed the standard error using ten Monte Carlo simulations with different seeds. Typically, the Monte Carlo convergence is expected to be $1/\sqrt{N_p}$, where N_p is the

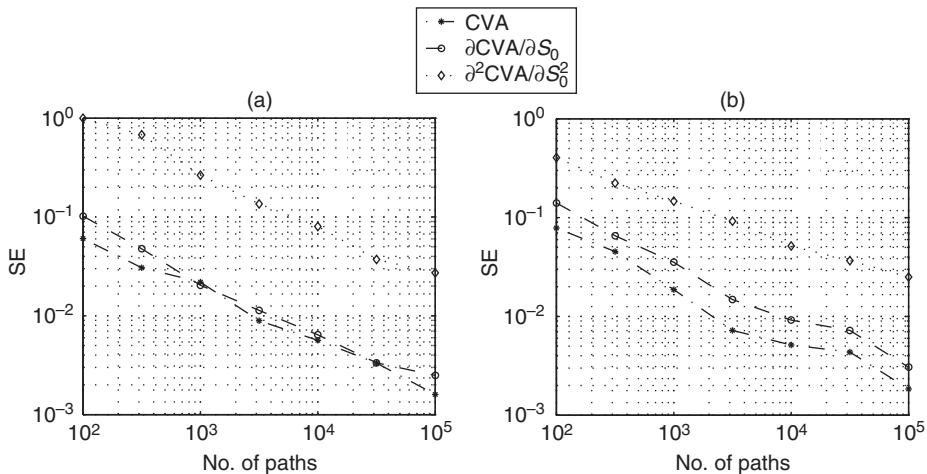
TABLE 3 Relative L_2 and L_∞ errors compared with the COS method using the linear or spline interpolation.

Error	Quantity	Linear interpolation		Spline interpolation	
		Test A	Test B	Test A	Test B
$\ \cdot\ _\infty$	EE	2.1728×10^{-3}	3.7094×10^{-3}	2.1690×10^{-3}	3.7006×10^{-3}
	PFE _{97.5%}	4.9120×10^{-3}	5.3274×10^{-3}	4.9213×10^{-3}	5.2280×10^{-3}
	$\frac{\partial \text{EE}}{\partial S_0}$	4.7905×10^{-3}	6.6173×10^{-3}	4.8058×10^{-3}	6.6243×10^{-3}
	$\frac{\partial^2 \text{EE}}{\partial S_0^2}$	3.5990×10^{-2}	3.7086×10^{-2}	3.5982×10^{-2}	3.7107×10^{-2}
	EE	1.8225×10^{-3}	3.1294×10^{-3}	1.8197×10^{-3}	3.1216×10^{-3}
	PFE _{97.5%}	3.0751×10^{-3}	5.2244×10^{-3}	3.0832×10^{-3}	5.2280×10^{-3}
$\ \cdot\ _2$	$\frac{\partial \text{EE}}{\partial S_0}$	3.8470×10^{-3}	5.3315×10^{-3}	3.8508×10^{-3}	5.2280×10^{-3}
	$\frac{\partial^2 \text{EE}}{\partial S_0^2}$	2.5988×10^{-2}	2.2843×10^{-2}	2.5921×10^{-2}	2.2881×10^{-2}

FIGURE 3 Error convergence of EE and first- and second-order sensitivities for tests A and B by increasing the number of grid points ($2m \times m$) used in the finite difference computation.

We use m in the v -direction and $2m$ in the s -direction. For every exposure computation, $\times 10^5$ paths are used, simulated with a fixed seed to avoid noise. Here, a spline interpolation is used, but similar analysis performed using a linear interpolation is shown in Table 3. (a) Test A. (b) Test B.

FIGURE 4 Convergence of the relative standard error (SE), Delta and Gamma of CVA for tests A and B for an increasing number of paths.



Here, the number of finite difference grid points is set equal to 350 in the v -direction and 700 in the s -direction, and the standard error is computed relative to the mean. (a) Test A. (b) Test B.

number of Monte Carlo paths. We see that, for both tests, all quantities converge as expected.

Note that the computation time of the FDMC method heavily depends on the number of grid points. When we compare the computation time of the price grids computed by COS or FDMC, we see that the FDMC method is significantly faster.⁶

4.1.3 Portfolio of options

For the evaluation of CVA, we assume a recovery rate of 40%. The hazard rate is computed by assuming a five-year CDS with a spread of 400bps paid quarterly. The euro discount factors are taken from April 2014, and the resulting survival probabilities up to one year are obtained as explained in Section 3.3. Because we assume the absence of wrong- and right-way risk, we can compute the CVA for any CDS spread. In this case, CVA is a linear function of the CDS spread.

Different options in one portfolio can have different strikes and maturities. Due to these different maturities, the finite difference procedure is faced with a discontinuity in time. To assess the possible effect on the accuracy, we consider a portfolio of two

⁶In our implementation, the COS grid was obtained by subsequently pricing every grid point, which probably can be optimized.

TABLE 4 Tested portfolios.

		Type	Maturity	Strike	Barrier
Portfolio I	Option 1	Call	$T_1 = 1$	$K_1 = 133$	—
	Option 2	Put	$T_2 = 0.4$	$K_2 = 138$	—
	Option 3	Barrier	$T_3 = 0.8$	$K_3 = 135$	$B_3 = 120$
Portfolio II	Option 1	Call	$T_1 = 1$	$K_1 = 133$	—
	Option 2	Put	$T_2 = 0.4$	$K_2 = 138$	—

European options with different strike and maturity. We again compare the resulting exposure profiles and CVA values with the Monte Carlo COS method. For the benchmark, we compute separate exposure profiles for every option with the Monte Carlo COS method and compute the EE of the portfolios as the sum. This is similar for the sensitivities that are obtained by a bump-and-revalue procedure per option. In the FDMC method, the call and put options are computed simultaneously on one grid. The barrier option is computed on a separate grid, because for every path termination needs to be checked. The resulting option prices per path are added to the portfolio values, and from this the mean and quantiles can be calculated.

Again, we assume that the Heston dynamics drive the underlying risk factors. The Heston parameters that drive the underlying are chosen as in case B of the previous subsection. All the options in the portfolio are written on this single FX rate. We consider two portfolios: portfolio I consists of a call, put and barrier option, while portfolio II consists only of a call and put. Table 4 shows the option parameters for the two portfolios.

The results presented in this paper also hold for portfolios consisting of an arbitrary larger number of options, but for illustrative reasons we present results for only three options.

In Table 5, we show the CVA values. Here, we computed the CVA as a percentage of the portfolio value. The sensitivities are quoted relative to the sensitivities of the initial portfolio. This way, we can quantify the change between the CVA-adjusted and the non-CVA-adjusted portfolio:

$$\text{CVA\%} := 100 \frac{\text{CVA}}{\Pi}, \quad (4.5)$$

$$\Delta_{S_0} := 100 \frac{\partial \text{CVA}}{\partial S_0} / \frac{\partial \Pi}{\partial S_0}, \quad (4.6)$$

$$\Gamma_{S_0} := 100 \frac{\partial^2 \text{CVA}}{\partial S_0^2} / \frac{\partial^2 \Pi}{\partial S_0^2}. \quad (4.7)$$

TABLE 5 CVA, Delta and Gamma for the portfolios I and II as a percentage of non-adjusted values.

	Linear interpolation		Spline interpolation	
	Portfolio I (%)	Portfolio II (%)	Portfolio I (%)	Portfolio II (%)
CVA FDMC	2.79	2.77	2.79	2.77
CVA COS	2.79	2.77	2.79	2.77
Δ_{S_0} splitting FDMC	23.49	18.78	23.29	18.72
Δ_{S_0} B&R COS	23.52	18.85	23.40	18.79
Γ_{S_0} splitting FDMC	2.58	2.42	2.57	2.41
Γ_{S_0} B&R COS	2.52	2.38	2.57	2.43

The percentages are computed by spline and linear interpolation for the FDMC method and the benchmark COS method. The sensitivities in the COS method are obtained by a bump-and-revalue technique.

By looking at Figures 5(a) and 5(b), we can see that the EE drops at $t = 0.4$, when the put option expires. This discontinuity is captured nicely by the FDMC method, where the put and call options are computed on one finite difference grid. By looking at Table 5, we can conclude that the resulting value adjustments are accurate compared with the Monte Carlo COS method. Next to that, the difference between spline and linear interpolation is small.

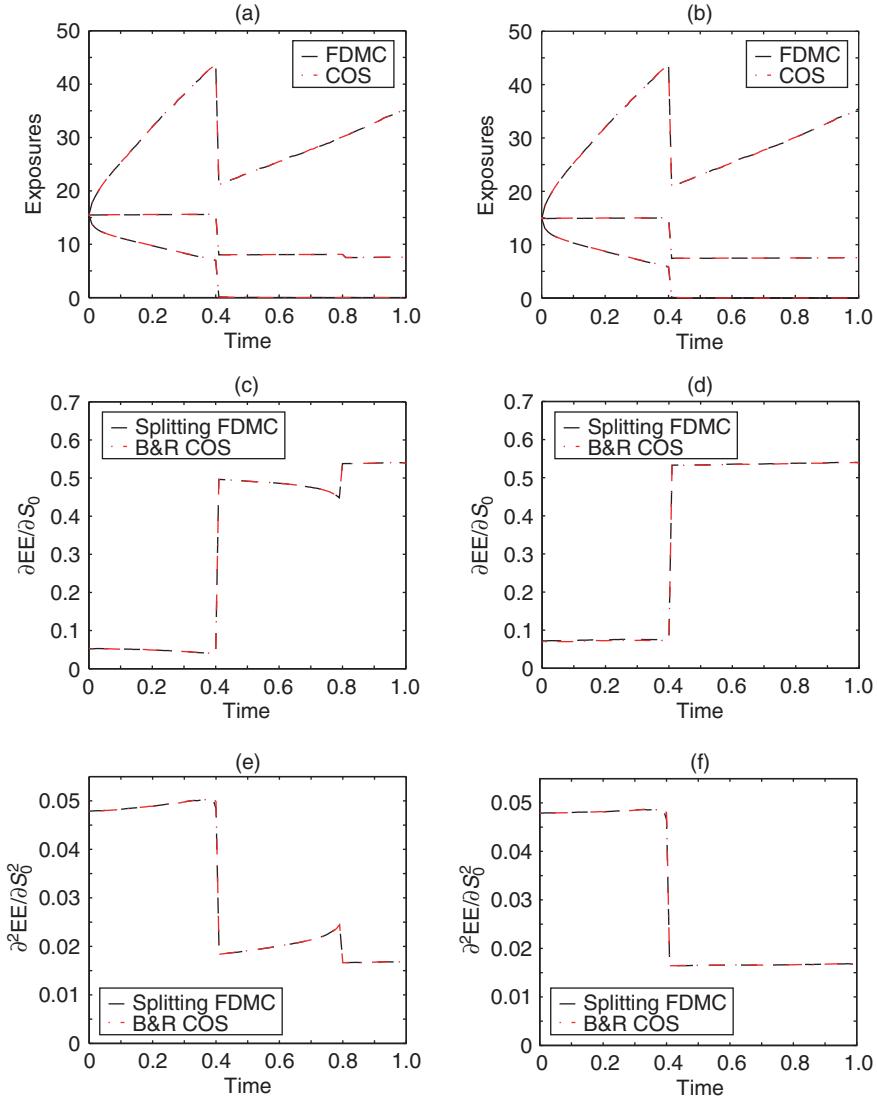
In portfolio I, the call and put options have a bigger effect on the EE than the barrier option. Also, the higher PFE is heavily affected by the expiry of the put option. If we compare Figures 5(a) and 5(b), we can see that the impact of the expiring barrier option at $T = 0.8$ is not reflected in the PFE, and it is only minor in the EE profile. This minor barrier effect is also visible when we compare the CVAs for portfolios I and II. The difference between these portfolios is due to the barrier option, and we can see a CVA difference of 1.5% in Table 5.

Next, if we look at the Delta profiles in Figures 5(c) and 5(d), we can see that for portfolio I the impact of the barrier option is reflected by a steep decrease at the expiry of the barrier option $T = 0.8$. As this barrier option is absent in portfolio II, this decrease is absent in the Delta profile of portfolio II. This impact is also confirmed by looking at the sensitivity of CVA with respect to S_0 in Table 5, where the difference between portfolios I and II is in the region of 25% for Δ_{S_0} .

In the case of Gamma, shown in Figures 5(e) and 5(f), the barrier option in portfolio I shows a steep increase at the expiry of the barrier option. The relative impact for Gamma, however, is smaller than for Delta. In Table 5, we see that the difference in Gamma between portfolios I and II is in the range of 3%.

Further, we can see that the spline interpolation yields similar results to the linear interpolation. These results indicate that the effect of barrier options in a portfolio can

FIGURE 5 Exposure, Delta and Gamma profiles for portfolios I and II over time for case B, computed with the FDMC or COS method.



Again, the COS sensitivities are computed running two or three simulations from a bumped initial value. The results compared with the COS method are accurate up to an order of $\times 10^{-3}$. (a) Exposure profiles for portfolio I. (b) Exposure profiles for portfolio II. (c) Delta profile for portfolio I. (d) Delta profile for portfolio II. (e) Gamma profile for portfolio I. (f) Gamma profile for portfolio II.

be more severe in the sense of sensitivities than in CVA itself. Clearly, a small change in the EE profile can have a bigger impact on the first- and second-order sensitivity. Further, in the online appendix, we show that for barrier options the sensitivities are also more sensitive to changes in moneyness levels.

4.2 Heston Hull–White model

As the parameters are not calibrated in this research, the mean-reverting level in the short rate process $\theta(t)$ is assumed to be constant over time.⁷

4.2.1 One-touch options

The OT option only delivers a fixed payoff at maturity; therefore, it has a discontinuous profile during the entire lifetime of the option. Next to that, the three-dimensional dynamics of the Heston Hull–White model implies a coarser grid in every dimension. Because of this discontinuity and the coarser grid, the interpolation scheme is more important. In Figure 6, we show the exposures for tests A and B. We can see that the 97.5% PFE in test B reaches the maximum payout level earlier than in test A. The lower mean-reversion speed in combination with the higher volatility of volatility and long-term mean of the variance process in test B causes fatter tails, which imply a higher hitting probability.

The spline and linear exposures are very close. This is further shown in Table 6, where we see that the L_∞ and L_2 differences are in the range of 1%.

4.2.2 Other sensitivities

Here, we look at Vega and Rho for OT options over time. Note that, in this case, the bump-and-revalue method uses two Monte Carlo simulations and estimates the derivative by

$$\frac{\partial \text{EE}(t)}{\partial \Theta} \approx \frac{\text{EE}_{\Theta+\epsilon_\Theta}(t) - \text{EE}_\Theta(t)}{\epsilon_\Theta}, \quad (4.8)$$

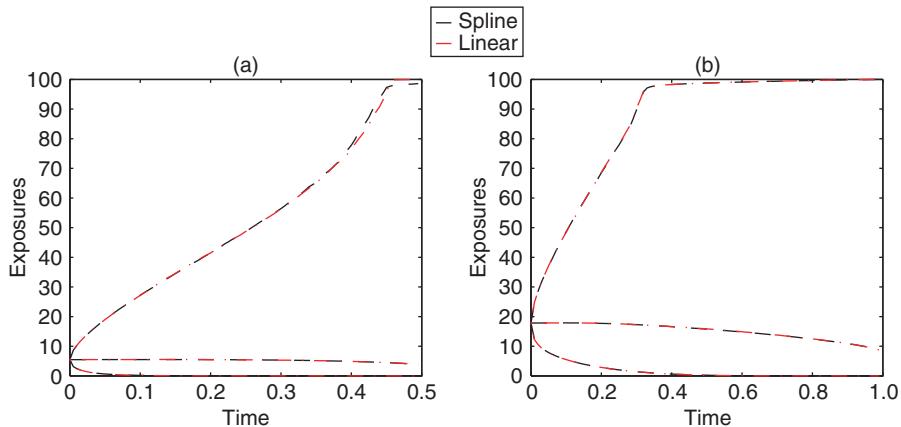
which is only first-order accurate in ϵ_Θ , which we choose as $0.01 \times V_0$ and $0.01 \times R_0^d$, respectively. Figure 7 shows that the sensitivities computed by the splitting scheme and the bump-and-revalue method agree over time. This is further confirmed by looking at the relative differences that are again in the range of 1%, as presented in Table 7.

5 CONCLUSION

In this research, we proposed a new computational technique to compute exposure profiles and their sensitivities. This paper extends the FDMC method described in

⁷A study on real model impact on exposure, in which parameters are calibrated to real market data, is a subject for future study.

FIGURE 6 Exposures and PFE for a one-year OT option with barrier level $1.2S_0$ and payout 100.

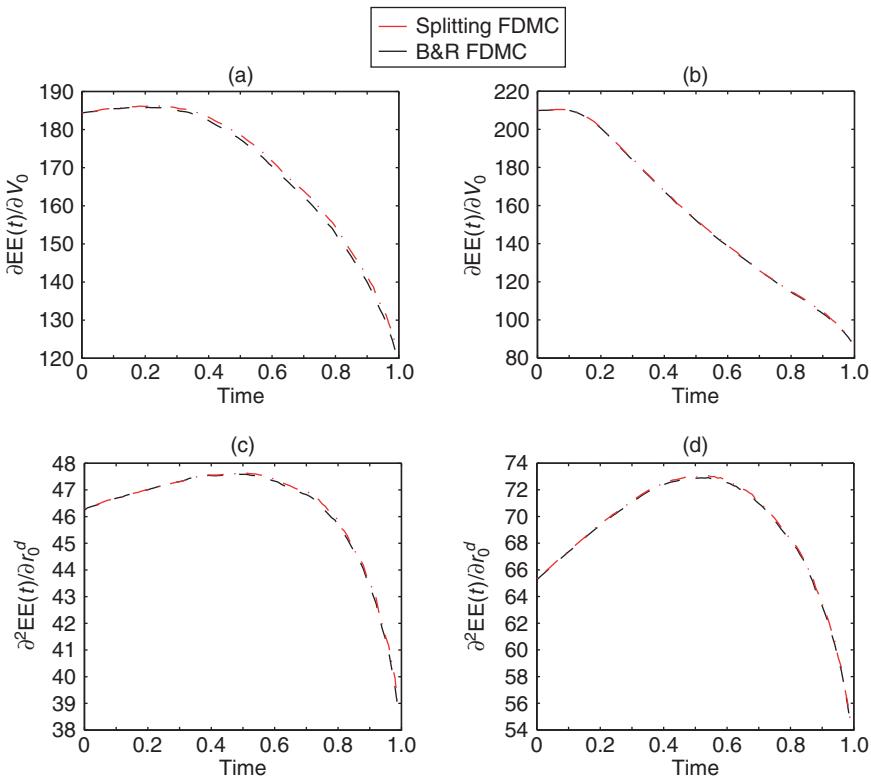


Here, we use 100 time points and the number of finite difference grid points is set equal to 100 in the s -direction and 50 in the v - and r^d -directions. (a) Test A. (b) Test B.

TABLE 6 Relative L_2 and L_∞ differences between exposure metrics, computed using linear and spline interpolation for one-year OT option, with barrier level $1.2S_0$ and payout 100.

Differences	Quantity	Test A	Test B
$\ \cdot\ _\infty$	EE	5.889×10^{-3}	5.819×10^{-3}
	PFE _{97.5%}	1.123×10^{-2}	6.310×10^{-3}
	PFE _{2.5%}	4.327×10^{-4}	1.066×10^{-3}
$\ \cdot\ _2$	EE	2.411×10^{-3}	3.002×10^{-3}
	PFE _{97.5%}	6.454×10^{-3}	1.410×10^{-3}
	PFE _{2.5%}	7.700×10^{-4}	1.504×10^{-3}

de Graaf *et al* (2014) and is based on combining the Monte Carlo scenario generation with option valuation by solving a PDE on a grid. For every scenario at every time point, the option prices are obtained by interpolating the scenarios on this option grid. The EE needed for the computation of CVA is computed by averaging. For the Heston model, we have shown that, compared with a benchmark Monte Carlo COS method, the FDMC is a computationally efficient and accurate method; it can therefore serve as an alternative to the widely used American-style Monte Carlo approach, which in application to exotic options can suffer from regression bias.

FIGURE 7 Derivatives of EE over time for OT options.

The derivatives are with respect to S_0 , V_0 and r_0^d . The sensitivities are computed by the splitting method and the bump-and-revalue method. We use 100 time points, and the number of finite difference grid points is set equal to 100 in the s -direction and 50 in the v - and r^d directions. (a) Vega profile for test A. (b) Vega profile for test B. (c) Rho profile for test A. (d) Rho profile for test B.

The sensitivities with respect to S_0 and R_0^d are obtained efficiently by leveraging from the finite difference grid. Compared with a “brute force” bump-and-revalue technique, the sensitivity results are accurate, and no extra Monte Carlo simulations are needed, which is a computational advantage. In the case of sensitivity with respect to V_0 , we show that when an analytic expression of the future variance with respect to initial variance is available, the same technique can be applied.

Under the Heston dynamics, we analyzed the accuracy of the method by comparing it with a benchmark solution, and we assess the convergence of the solutions by (1) increasing the number of paths in the Monte Carlo simulation and (2) increasing the number of grid points used in the finite difference procedure. As expected, the

TABLE 7 Relative L_2 and L_∞ differences between future exposure sensitivities, computed using a splitting scheme and a bump-and-revalue technique for one-year OT option, with barrier level $1.2S_0$ and payout 100.

Difference	Quantity	Test A	Test B
$\ \cdot\ _\infty$	$\frac{\partial \text{EE}}{\partial V_0}$	1.141×10^{-2}	4.325×10^{-3}
	$\frac{\partial \text{EE}}{\partial R_0^d}$	6.872×10^{-3}	3.922×10^{-3}
$\ \cdot\ _2$	$\frac{\partial \text{EE}}{\partial V_0}$	6.582×10^{-3}	2.478×10^{-3}
	$\frac{\partial \text{EE}}{\partial R_0^d}$	1.946×10^{-3}	1.813×10^{-3}

standard error converges by $1/\sqrt{N}$, where N is the number of paths. By increasing the number of grid points, the relative error converges in first order.

Next, we showed that we can use the method to compute exposure profiles for a portfolio of options with different maturities. In this portfolio, the EEs of all options that are not path-dependent (European options) can be efficiently computed on a single grid. The resulting discontinuity in time is captured, and no significant error propagation is observed. The EEs for path-dependent options have to be computed individually and are added to the portfolio before computing the means. The sensitivities can again be computed with small extra computational time. Results compared with the Monte Carlo COS method are accurate for both linear and spline interpolation.

To further assess the impact of the interpolation, OT options that have a discontinuity over time are considered. In combination with the computationally challenging Heston Hull–White model, where less grid points can be used, the interpolation is essential. We found that even in this case a linear interpolation is sufficient, as differences are smaller than 1.2% for both exposures and sensitivities.

In forthcoming research, we assess in detail the effect of skew and stochastic interest rate on CVA and its sensitivities by using model parameters calibrated to real market data and a wide range of option contract parameters.

DECLARATION OF INTEREST

The authors report no conflict of interest. The authors alone are responsible for the writing of this work. This work was supported by the Dutch Technology Foundation STW under project 12214. Peter Sloot acknowledges support from the Russian Scientific foundation, Grant 14-21-00137.

ACKNOWLEDGEMENTS

We would like to thank Qian Feng for helpful advice regarding the Monte Carlo COS method. Further, we thank Norbert Hari, Shashi Jain and Sarunas Simaitis for useful discussions about the FDMC method.

REFERENCES

- Albrecher, H., Mayer, P., Schoutens, W., and Tistaert, J. (2007). The little Heston trap. *Wilmott Magazine*, January, 83–92.
- Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model. *The Journal of Computational Finance* **11**(3), 1–42 (<http://doi.org/bm4f>).
- Basel Committee on Banking Supervision (2010). Basel III: a global regulatory framework for more resilient banks and banking systems. Technical Report, June.
- Broadie, M., and Glasserman, P. (1996). Estimating security price derivatives using simulation. *Management Science* **42**(2), 269–285 (<http://doi.org/bmjd96>).
- Chan, J. H., and Joshi, M. (2010). First and second order Greeks in the Heston model. Working Paper 1718102, Social Science Research Network (<http://doi.org/fzp474>).
- de Graaf, C. S. L., Feng, Q., Kandhai, B. D., and Oosterlee, C. W. (2014). Efficient computation of exposure profiles for counterparty credit risk. *International Journal of Theoretical and Applied Finance* **17**(4), 1–24 (<http://doi.org/bm4g>).
- Ekström, E., and Tysk, J. (2011). Boundary conditions for the single-factor term structure equation. *Annals of Applied Probability* **21**(1), 332–350 (<http://doi.org/cwhvxx>).
- Fang, F., and Oosterlee, C. W. (2011). A Fourier-based valuation method for Bermudan and barrier options under Heston's model. *SIAM Journal on Financial Mathematics* **2**(1), 439–463 (<http://doi.org/b2pshv>).
- Feng, Q., and Oosterlee, C. W. (2014). Monte Carlo calculation of exposure profiles and Greeks for Bermudan and barrier options under the Heston Hull–White model. Working Paper 2494233, Social Science Research Network (<http://doi.org/bm4h>).
- Gregory, J. (2010). *Counterparty Credit Risk*. Wiley Online Library.
- Guo, S., Grzelak, L. A., and Oosterlee, C. W. (2013). Analysis of an affine version of the Heston Hull–White option pricing partial differential equation. *Applied Numerical Mathematics* **72**(October), 143–159 (<http://doi.org/bm4j>).
- Haentjens, T., and In't Hout, K. J. (2012). ADI finite difference schemes for the Heston Hull–White PDE. *The Journal of Computational Finance* **16**(1), 83–110 (<http://doi.org/bm4k>).
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**(2), 327–343 (<http://doi.org/fg525s>).
- Hull, J., and White, A. (1993). One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis* **28**(2), 235–254 (<http://doi.org/bkbdm6>).
- Hull, J., and White, A. (2012). CVA and wrong-way risk. *Financial Analysts Journal* **68**(5), 58–69 (<http://doi.org/bm4m>).
- Hundsdorfer, W., and Verwer, J. G. (2003). *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Volume 33. Springer Series in Computational Mathematics. Springer (<http://doi.org/bm4n>).

- In't Hout, K. J., and Foulon, S. (2010). ADI finite difference schemes for option pricing. *International Journal of Numerical Analysis and Modeling* **7**(2), 303–320.
- In't Hout, K. J., and Welfert, B. D. (2009). Unconditional stability of second-order ADI schemes applied to multi-dimensional diffusion equations with mixed derivative terms. *Applied Numerical Mathematics* **59**(3-4), 677–692 (<http://doi.org/dg3hrh>).
- Ng, L., and Peterson, D. (2009). Potential future exposure calculations using the BGM model. *Wilmott Journal* **1**(4), 213–225 (<http://doi.org/c8ptbz>).
- Ng, L., Peterson, D., and Rodriguez, A. E. (2010). Potential future exposure calculations of multi-asset exotic products using the stochastic mesh method. *The Journal of Computational Finance* **14**(2), 119–153 (<http://doi.org/bm4p>).
- Schoutens, W., Simons, E., and Tistaert, J. (2004). A perfect calibration! Now what? *Wilmott Magazine*, March, 66–78.
- Shen, Y., Weide, J. A. M., and Anderluh, J. H. M. (2013). A benchmark approach of counterparty credit exposure of Bermudan option under Lévy process: the Monte Carlo–COS method. *Procedia Computer Science* **18**, 1163–1171 (<http://doi.org/bm4q>).
- Tavella, D., and Randall, C. (2000). *Pricing Financial Instruments: The Finite Difference Method*. Wiley.
- Whetten, M., Adelson, M., and Bemmelen, M. V. (2004). Credit default swap (CDS) primer. *Nomura Fixed Income Research*, April, 1–12.

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net



Research Paper

Smile with the Gaussian term structure model

Abdelkoddousse Ahdida, Aurélien Alfonsi and Ernesto Palidda

Université Paris-Est, CERMICS, Projet MATHRISK ENPC-INRIA-UMLV,
6–8 Avenue Blaise Pascal, 77455 Marne La Vallée, France;
emails: ahdida.abdel@gmail.com, alfonsi@cermics.enpc.fr, ernesto.palidda@gmail.com

(Received December 22, 2014; revised November 4, 2015; accepted December 2, 2015)

ABSTRACT

We propose an affine extension of the linear Gaussian term structure model (LGM) such that the instantaneous covariation of the factors is given by an affine process on semidefinite positive matrixes. We begin by setting up the model and presenting some important properties concerning the Laplace transform of the factors and the ergodicity of the model. Then, we present two main numerical tools for implementing the model in practice. First, we obtain an expansion of caplet and swaption prices around the LGM. Such a fast and accurate approximation is useful in assessing model behavior on the implied volatility smile. Second, we provide a second-order scheme for the weak error, which enables us to calculate exotic options by a Monte Carlo algorithm. These two pricing methods are compared with the standard method based on Fourier inversion.

Keywords: affine term structure model (ATSM); linear Gaussian model (LGM); Wishart processes; price expansion; discretization scheme; swaptions.

1 MOTIVATION AND OVERVIEW OF THE PAPER

Affine term structure models (ATSMs) are an important class of models for interest rates that include the classical and pioneering models of Vasicek (1977) and Cox *et al* (1985). These models have been settled and popularized by the papers of Duffie and Kan (1996), Dai and Singleton (2000) and Duffie *et al* (2003). We refer the reader to Filipović (2009) for a textbook on these term structure models. The linear Gaussian model (LGM) is a simple but important subclass of ATSMs, which assumes that underlying factors follow a Gaussian process. The model has been considered by El Karoui *et al* (1991) and El Karoui and Lacoste (1992), and it has now become a market standard for pricing fixed income derivatives, thanks to its simplicity. However, this model has one main drawback when calibrated to market data: it produces implied volatility smiles that are flat.

The goal of this paper is to present a quite natural extension of the LGM that keeps the affine structure and generates an implied volatility smile. To do so, we consider a Wishart-type affine diffusion on a set of semidefinite positive matrixes, and replace, roughly speaking, the constant volatility matrix by (a linear function of) this process. The dependence between the factors and their volatility is created via a specific covariation that keeps the affine structure; this has been proposed by Da Fonseca *et al* (2008) in an equity framework. Thanks to this, the proposed model, which is a stochastic variance–covariance affine term structure model (see Definition 3.6), is able to produce an implied volatility smile. It has many parameters and may, at first glance, seem difficult to handle. For this reason, we present it as a perturbation of the LGM. Thus, the calibration of the model to market data can be completed in two steps. First, we calibrate the LGM, and then we calibrate the new parameters to the implied volatility smile. The calibration of this model is discussed for some cases in Palidda (2015). In the present paper, we do not tackle the practical calibration issue: our goal is just to set up the model and give the main numerical methods for its practical use. In Section 3, we define the model and present some of its important properties, such as the value of the Laplace transform under the initial and forward measures and the ergodicity property. Then, we present two tools that are important in implementing the model in practice. First, in Section 4, we present price expansions for caplets and swaptions around the LGM when the volatility of the volatility of the factor Y is small. These explicit formulas are useful for quickly calculating the impact of the parameters on the volatility cube, and thus for calibrating the model. Second, in Section 5, we propose a discretization scheme for the model that is of second order for the weak error. Having an accurate scheme is important in practice, since it allows us to calculate exotic options via a Monte Carlo algorithm. In addition, this scheme can be easily adapted to other models relying on the same affine structure, such as that of Da Fonseca *et al* (2008). Finally, Section 6 compares the expansion and the

Monte Carlo method with the classical Fourier technique popularized by Carr and Madan (1999), and indicates the relevance of each method.

2 THE LINEAR GAUSSIAN MODEL IN A NUTSHELL

The model that we present is meant to extend the classical LGM; thus, we need to briefly recall the LGM. We work under a risk-neutral measure \mathbb{P} , and consider a p -dimensional standard Brownian motion Z . Let Y be the solution of the following Ornstein–Uhlenbeck stochastic differential equation (SDE):

$$Y_t = y + \int_0^t \kappa(\theta - Y_s) ds + \int_0^t \sqrt{V} dZ_s, \quad (2.1)$$

where $\kappa \in \mathcal{M}_p(\mathbb{R})$ is a matrix of order p , V is a semidefinite positive matrix of order p and $\theta \in \mathbb{R}^p$. The LGM assumes that the spot rate is an affine function of the vector Y ,

$$r_t = \varphi + \sum_{i=1}^p Y_t^i, \quad (2.2)$$

and the coordinates Y^i are usually called the factors of the model. It is not restrictive to assume that the weight of each factor in (2.2) is the same for all factors and equal to 1. If we had $r_t = \varphi + \sum_{i=1}^p m_i Y_t^i$, we could easily check that $(m_1 Y^1, \dots, m_p Y^p)^T$ is also an Ornstein–Uhlenbeck process. ATSMs generally assume that the parameters (here, κ , θ and V) are fixed and valid over a long time period, while the factors (here, the vector Y) evolve and reflect the current state of the market. Therefore, one often assumes that the process Y is stationary in order to reflect some market equilibrium. Also, the factors are usually associated with different time scales: a factor with a small (respectively, large) mean reversion will influence the long-term (respectively, short-term) behavior of the interest rate. This leads us to assume that

$$\kappa = \text{diag}(\kappa_1, \dots, \kappa_p), \quad \text{with } 0 < \kappa_1 < \dots < \kappa_p.$$

We work under this assumption in what follows. It can easily be checked (see, for example, Andersen and Piterbarg 2010) that any LGM such that κ has distinct positive eigenvalues can be rewritten, up to a linear transformation of the factors, within the present parameterization.

Let $(\mathcal{F}_t)_{t \geq 0}$ denote the natural filtration of Z . For $0 \leq t \leq T$, the price $P_{t,T} = \mathbb{E}[\exp(-\int_t^T r_s ds) | \mathcal{F}_t]$ at time t of the zero-coupon bond with maturity T is an exponential affine function of Y :

$$P_{t,T} = \exp(E(T-t) + B(T-t)^T Y_t), \quad (2.3)$$

where $B(\tau) = -(\kappa^T)^{-1}(I_p - e^{-\kappa^T \tau})\mathbf{1}_p$ and

$$E(\tau) = -\varphi\tau + \int_0^\tau B(s)^T \kappa \theta + (B(s)^T V B(s)/2) ds$$

for $\tau \geq 0$. Here, $\mathbf{1}_p$ stands for the vector in \mathbb{R}^p that has all its entries equal to 1. The function $B(\tau)$ maps the factors' variations ΔY on the yield curve variations and is often called the support function. The factors Y^i associated with the larger parameters κ_i affect the short-term behavior of the yield curve, while those factors associated with the smaller parameters κ_i will drive the long-term behavior.

We now briefly introduce some of the basic notions on the interest rate vanilla option market. The most liquid traded interest rate options are swaptions and caplets. These are expressed with respect to the forward London interbank offered rate (Libor) and the forward swap rate, which are defined as follows for $0 \leq t \leq T$, $\delta > 0$ and $m \in \mathbb{N}^*$:

$$L_t(T, \delta) = \frac{1}{\delta} \left(\frac{P_{t,T}}{P_{t,T+\delta}} - 1 \right),$$

$$S_t(T, m) = \frac{P_{t,T} - P_{t,T+m\delta}}{\delta \sum_{i=1}^m P_{t,T+i\delta}}.$$

The respective prices of caplets and swaptions are given by

$$C_t(T, \delta, K) = \mathbb{E}[\exp \left(- \int_t^{T+\delta} r_s ds \right) (L_T(T, \delta) - K)^+ | \mathcal{F}_t],$$

Swaption _{t} (T, m, δ, K)

$$= \mathbb{E} \left[\exp \left(- \int_t^T r_s ds \right) \sum_{i=1}^m \delta P_{t,T+i\delta} (S_T(T, m) - K)^+ \mid \mathcal{F}_t \right].$$

Caplets are usually available for short tenors δ (up to one year), and swaptions are quoted for tenors $m\delta$ (from two to thirty years). The market practice is to apply a standard change of numeraire technique (see Geman *et al* 1995) and rewrite the above expressions as

$$C_t(T, \delta, K) = P_{t,T+\delta} \mathbb{E}^{T+\delta} [(L_T(T, \delta) - K)^+ | \mathcal{F}_t], \quad (2.4)$$

$$\text{Swaption}_t(T, m, \delta, K) = \left(\sum_{i=1}^m \delta P_{t,T+i\delta} \right) \mathbb{E}^A [(S_T(T, m) - K)^+ | \mathcal{F}_t], \quad (2.5)$$

where $\mathbb{E}^{T+\delta}$ (respectively, \mathbb{E}^A) denotes the expectation taken with respect to the measure $T + \delta$ -forward neutral (respectively, annuity) measure associated with the numeraire $P_{t,T+\delta}$ (respectively, $\sum_{i=1}^m \delta P_{t,T+i\delta}$). The market prices are then quoted

and analyzed in terms of either the lognormal or normal implied volatility obtained by inverting the pricing formulas (2.4) and (2.5) with respect to the Black–Scholes and Bachelier formulas. Within the LGM model, the lognormal implied volatility of the caplet is given by

$$\int_t^T [B(T-u) - B(T+\delta-u)]^T V [B(T-u) - B(T+\delta-u)] du,$$

which is a particular case of (4.4) below. This implied volatility does not depend on the strike. It shows that the mean-reversion parameter κ plays a role in shaping the form of the caplet volatility cube, according to the different time scales. The role of the diagonal coefficients of the matrix V is determined by the support functions

$$m_{ii}(\tau, \delta) = \left(\frac{1 - e^{-\kappa_i \delta}}{\kappa_i} \right)^2 \frac{1 - e^{-2\kappa_i \tau}}{2\kappa_i \tau}.$$

The effect of the off-diagonal elements of the variance–covariance matrix V is determined by the support functions

$$m_{ij}(\tau, \delta) = \frac{1 - e^{-\kappa_i \delta}}{\kappa_i} \frac{1 - e^{-\kappa_j \delta}}{\kappa_j} \frac{1 - e^{-(\kappa_i + \kappa_j)\tau}}{(\kappa_i + \kappa_j)\tau}.$$

These functions are plotted in Figure 1.

Also, by using a standard approximation, we can obtain the normal implied volatility of the swaptions:

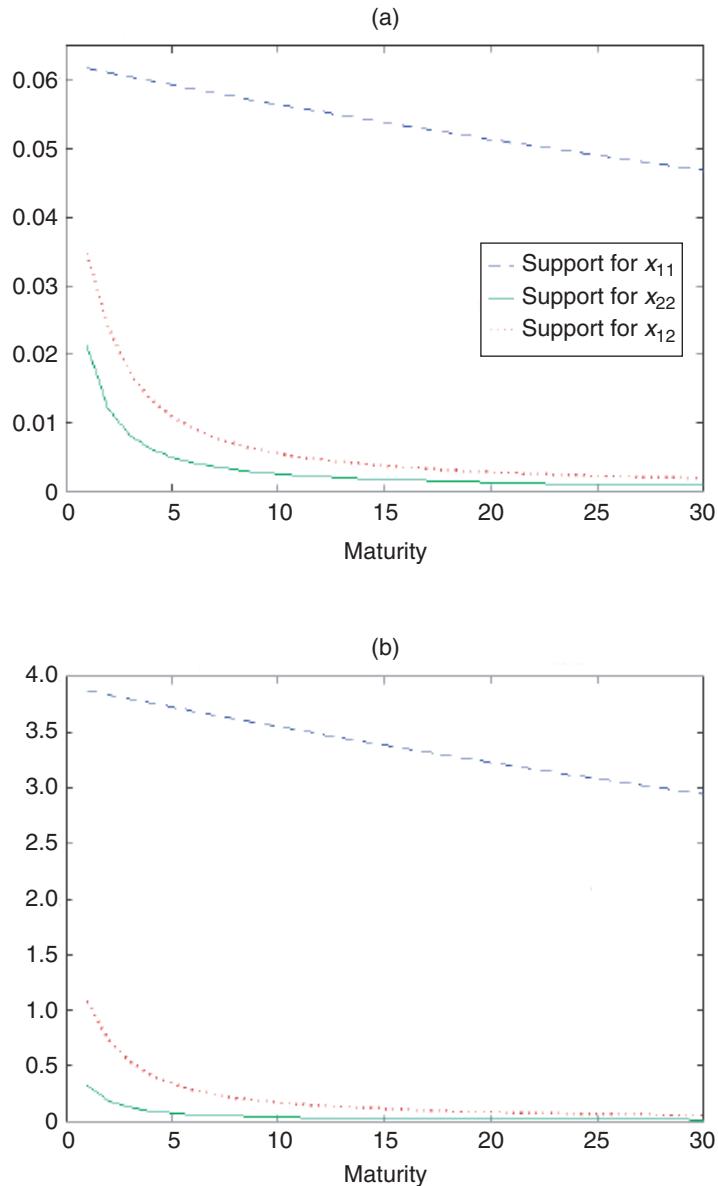
$$\int_t^T [B^S(u)]^T V B^S(u) du,$$

with

$$\begin{aligned} B^S(u) &= \omega_0^0 B(T-u) - \omega_0^m B(T+m\delta-u) \\ &\quad - S_0(T, m, \delta) \sum_{k=1}^m \omega_0^k B(T+k\delta-u), \\ \omega_0^k &= \frac{P_{0,T+k\delta}}{\sum_{i=1}^m P_{0,T+i\delta}}. \end{aligned}$$

This is a particular case of (4.19) below. This implied volatility has a rather similar structure to that of the caplets, but it is not time homogeneous. The implied volatilities for caplets and swaptions do not depend on the strike and thus give a flat smile, which is a well-known fact. This is unfortunate if one is aiming to reproduce the volatility cube observed on market data (ie, the implied volatility with respect to the maturity T , the tenor δ or $m\delta$ and the strike K). The extension of the LGM that we introduce in Section 3 is meant to correct this drawback.

FIGURE 1 Support functions for the volatility term structure in a two-factor model, with $\kappa = \text{diag}(0.01, 1)$ for (a) three-month and (b) two-year maturity.



3 AN AFFINE EXTENSION OF THE LINEAR GAUSSIAN MODEL WITH STOCHASTIC COVARIANCE

This section is devoted to the definition of the model that we study in this paper. This model is a stochastic variance–covariance perturbation of the LGM. We chose a quite general specification that keeps the model affine and gives a stochastic instantaneous covariance for the factors, which will generate a smile for the caplets and swaptions. We first present the dynamics of the factor and then present some properties of the model that rely on its affine structure.

3.1 State variables dynamics

We consider W , a d -by- d square matrix made of independent standard Brownian motions, and Z , an independent Brownian motion of dimension p . We will denote by $(\mathcal{F}_t)_{t \geq 0}$ the filtration generated by (W, Z) . We consider the following SDE for the state variables (or factors):

$$Y_t = y + \int_0^t \kappa(\theta - Y_s) ds + \int_0^t c \sqrt{X_s} [\bar{\rho} dZ_s + dW_s \rho], \quad (3.1)$$

$$\begin{aligned} X_t = x + \int_0^t (\Omega + (d-1)\epsilon^2 I_d^n + bX_s + X_s b^T) ds \\ + \epsilon \int_0^t \sqrt{X_s} dW_s I_d^n + I_d^n dW_s^T \sqrt{X_s}. \end{aligned} \quad (3.2)$$

The matrix I_d^n is defined for $0 \leq n \leq d$ by $(I_d^n)_{i,j} = \mathbb{1}_{i=j \leq n}$, and the parameters are taken as follows:

$$\begin{aligned} x, \Omega &\in \mathcal{S}_d^+(\mathbb{R}), b \in \mathcal{M}_d(\mathbb{R}), \epsilon \in \mathbb{R}_+, y, \theta \in \mathbb{R}^p, \\ \kappa &= \text{diag}(\kappa_1, \dots, \kappa_p) \quad \text{with } \kappa_1, \dots, \kappa_p > 0, \\ c \in \mathcal{M}_{p \times d}(\mathbb{R}), \rho &\in \mathbb{R}^d \quad \text{such that } |\rho|^2 := \sum_{i=1}^d \rho_i^2 \leq 1 \text{ and } \bar{\rho} = \sqrt{1 - |\rho|^2}, \end{aligned} \quad (3.3)$$

where $\mathcal{S}_d^+(\mathbb{R})$, $\mathcal{M}_d(\mathbb{R})$ and $\mathcal{M}_{p \times d}(\mathbb{R})$ denote, respectively, the set of semidefinite positive matrixes of order d , the set of square matrixes of order d and the set of matrixes with p rows and d columns. The process X is an affine diffusion on $\mathcal{S}_d^+(\mathbb{R})$, and the instantaneous covariance at time t of the factors Y is given by $c X_t c^T$. When $\epsilon = 0$ and $\Omega = -bx - xb^T$, we have $X_t = x$ and get back the Gaussian model with $V = cx c^T$. The dependence structure between Y and X through the driving Brownian motions is the same as that proposed by Da Fonseca *et al* (2008). As explained in Da Fonseca *et al* (2008), this is the most general way to get a nontrivial instantaneous correlation between Y and X while keeping the affine structure. In particular, the

instantaneous quadratic covariations are linear with respect to (Y, X) . We have, for $1 \leq i, j, k, l \leq d$ and $1 \leq m, m' \leq p$,

$$\langle d(Y_t)_m, d(Y_t)_{m'} \rangle = (c X_t c^T)_{m,m'} dt, \quad (3.4)$$

$$\begin{aligned} \langle d(X_t)_{i,j}, d(X_t)_{k,l} \rangle &= \epsilon^2 [(X_t)_{i,k} \mathbb{1}_{j=l \leq n} + (X_t)_{i,l} \mathbb{1}_{j=k \leq n} \\ &\quad + (X_t)_{j,k} \mathbb{1}_{i=l \leq n} + (X_t)_{j,l} \mathbb{1}_{i=k \leq n}] dt, \end{aligned} \quad (3.5)$$

$$\langle d(Y_t)_m, d(X_t)_{i,j} \rangle = \epsilon [(c X_t)_{m,i} (I_d^n \rho)_j + (c X_t)_{m,j} (I_d^n \rho)_i] dt. \quad (3.6)$$

We note that only the n first components of ρ matter, and we can assume without loss of generality that $\rho_{n+1} = \dots = \rho_d = 0$.

From (3.2), we easily get

$$e^{\kappa t} Y_t = y + \int_0^t e^{\kappa s} \kappa \theta ds + \int_0^t e^{\kappa s} c \sqrt{X_s} [\bar{\rho} dZ_s + dW_s \rho].$$

Therefore, the process Y is uniquely determined once the processes Z , W and X are given. We know by Cuchiero *et al* (2011) that the SDE on X has a unique weak solution when $x \in \mathcal{S}_d^+(\mathbb{R})$ and $\Omega \in \mathcal{S}_d^+(\mathbb{R})$, and a unique strong solution if we assume besides that x is invertible and $\Omega - 2\epsilon^2 I_d^n \in \mathcal{S}_d^+(\mathbb{R})$. This leads to the following result.

PROPOSITION 3.1 *If $x \in \mathcal{S}_d^+(\mathbb{R})$, $\Omega \in \mathcal{S}_d^+(\mathbb{R})$, there exists a unique weak solution of the SDE (3.2). If we assume, moreover, that $\Omega - 2\epsilon^2 I_d^n \in \mathcal{S}_d^+(\mathbb{R})$ and $x \in \mathcal{S}_d^+(\mathbb{R})$ is positive definite, there is a unique strong solution for the SDE (3.2).*

The affine structure of the process (X, Y) allows us to give formulas for the Laplace transform of the marginal laws by means of matrix Riccati differential equations (MRDEs). Similar calculations have been made in equity modeling by Da Fonseca *et al* (2008) and Benabid *et al* (2008). The following proposition states the precise result, which is useful for the pricing of zero-coupon bonds.

PROPOSITION 3.2 *Let $\Lambda, \bar{\Lambda} \in \mathbb{R}^p$ and $\Gamma, \bar{\Gamma} \in \mathcal{S}_d(\mathbb{R})$. For $t \geq 0$, we define $\lambda(t) \in \mathbb{R}^p$ by*

$$\lambda_i(t) = \Lambda_i e^{-\kappa_i t} + \frac{\bar{\Lambda}_i}{\kappa_i} (1 - e^{-\kappa_i t}). \quad (3.7)$$

Let us assume that there exists $\Upsilon \in \mathcal{S}_d(\mathbb{R})$ such that

$$\Upsilon - \Gamma \in \mathcal{S}_d^+(\mathbb{R}), \quad (3.8)$$

$$\begin{aligned} -[2\epsilon^2 \Upsilon I_d^n \Upsilon + \Upsilon (b + \frac{1}{2}\epsilon I_d^n \rho \lambda^T c) \\ + (b + \frac{1}{2}\epsilon I_d^n \rho \lambda^T c)^T \Upsilon + \frac{1}{2} c^T \lambda \lambda^T c + \bar{\Gamma}] \in \mathcal{S}_d^+(\mathbb{R}) \end{aligned} \quad (3.9)$$

for all $t \geq 0$. Then, the following system of differential equations

$$\begin{aligned} \dot{g} &= 2\epsilon^2 g I_d^n g + g(b + \frac{1}{2}\epsilon I_d^n \rho \lambda^T c) + (b + \frac{1}{2}\epsilon I_d^n \rho \lambda^T c)^T g \\ &\quad + \frac{1}{2} c^T \lambda \lambda^T c + \bar{\Gamma}, \quad g(0) = \Gamma, \end{aligned}$$

$$\dot{\eta} = \lambda^T \kappa \theta + \text{Tr}(g(\Omega + \epsilon^2(d-1)I_d^n)), \quad \eta(0) = 0, \quad (3.10)$$

has a unique solution, which is defined on \mathbb{R}_+ . It satisfies $\Upsilon - g(t) \in \mathcal{S}_d^+(\mathbb{R})$ for any $t \geq 0$. In addition, we have, for any $0 \leq t \leq T$,

$$\begin{aligned} \mathbb{E}\left[\exp\left(\text{Tr}(\Gamma X_T) + \Lambda^\top Y_T + \int_t^T \text{Tr}(\bar{\Gamma} X_s) + \bar{\Lambda}^\top Y_s ds\right) \middle| \mathcal{F}_t\right] \\ = \exp(\eta(T-t) + \text{Tr}(g(T-t)X_t) + \lambda(T-t)^\top Y_t). \end{aligned} \quad (3.11)$$

PROOF The proof is quite standard for affine diffusion. First, we note that if (3.11) holds, we necessarily have that

$$M_t = \exp\left(\int_0^t \text{Tr}(\bar{\Gamma} X_s) + \bar{\Lambda}^\top Y_s ds\right) \exp(\eta(T-t) + \text{Tr}(g(T-t)X_t) + \lambda(T-t)^\top Y_t)$$

is a martingale. We apply Itô's formula and use (3.4), (3.5) and (3.6). The martingale property yields

$$\begin{aligned} & \bar{\Gamma} X_t + \bar{\Lambda}^\top Y_t - \dot{\eta}(T-t) - \text{Tr}(\dot{g}(T-t)X_t) - \dot{\lambda}(T-t)^\top Y_t \\ & + \text{Tr}(g(T-t)[\Omega + (d-1)\epsilon^2 I_d^n + bX_s + X_s b^\top]) \\ & + \lambda(T-t)^\top \kappa(\theta - Y_t) + 2\epsilon^2 \text{Tr}(Xg(T-t)I_d^n g(T-t)) \\ & + \frac{1}{2} \text{Tr}(Xc^\top \lambda(T-t) \lambda^\top(T-t)c) \\ & + \frac{1}{2}\epsilon \text{Tr}(X[c^\top \lambda(T-t) \rho^\top I_d^n g(T-t) + g(T-t)I_d^n \rho \lambda^\top(T-t)c]) = 0. \end{aligned}$$

By identifying the constant term and the linear terms with respect to Y_t and X_t , we get (3.10) and $\dot{\lambda} = -\kappa\lambda + \bar{\Lambda}$, $\lambda(0) = \Lambda$; this leads to (3.7), since κ is diagonal with positive entries. By applying Dieci and Eirola (1994, Proposition 1.1) to $\Upsilon - g$, the solution of (3.10) exists and is well defined for $t \geq 0$.¹ In addition, $\Upsilon - g$ stays in $\mathcal{S}_d^+(\mathbb{R})$ by using (3.8) and (3.9).

It then remains to check that we indeed have (3.11), and it is sufficient to check it for $t = 0$. To do so, we apply Itô's formula to M and get

$$\begin{aligned} dM_s = M_s [\text{Tr}(g(T-s)[\sqrt{X_s} dW_s I_d^n + I_d^n dW_s^\top \sqrt{X_s}]) \\ + \lambda(T-s)^\top c \sqrt{X_s} [\bar{\rho} dZ_s + dW_s \rho]]. \end{aligned}$$

Thus, M is a positive local martingale (and, thus, a supermartingale), which gives $M_0 \geq \mathbb{E}[M_T]$. To prove that $M_0 = \mathbb{E}[M_T]$, we use the argument presented by Rydberg (1997). We define $N_t = M_t/M_0$ in order to work with probability measures. We define for $K > 0$, $\tau_K = \inf\{t \geq 0, \text{Tr}(X_t) \geq K\}$,

$$\pi_K(x) = \mathbb{1}_{\text{Tr}(x) \leq K} x + \mathbb{1}_{\text{Tr}(x) \geq K} \frac{K}{\text{Tr}(x)} x$$

¹ We thank Martino Grasselli for pointing out this reference to us.

for $x \in \mathcal{S}_d^+(\mathbb{R})$, and we consider $N_t^{(\mathbf{K})}$ the solution of

$$\begin{aligned} dN_s^{(\mathbf{K})} &= N_s^{(\mathbf{K})} \{ \text{Tr}(g(T-s)[\sqrt{\pi_{\mathbf{K}}(X_s)} dW_s I_d^n + I_d^n dW_s^T \sqrt{\pi_{\mathbf{K}}(X_s)}] \\ &\quad + \lambda(T-s)^T c \sqrt{\pi_{\mathbf{K}}(X_s)} [\bar{\rho} dZ_s + dW_s \rho] \}, \\ N_0^{(\mathbf{K})} &= 1. \end{aligned}$$

Clearly, $\mathbb{E}[N_T^{(\mathbf{K})}] = 1$, and under $d\mathbb{P}^{(\mathbf{K})}/d\mathbb{P} = N_T^{(\mathbf{K})}$,

$$dW_t^{(\mathbf{K})} = dW_t - 2\sqrt{\pi_{\mathbf{K}}(X_t)} g(T-t) I_d^n - \sqrt{\pi_{\mathbf{K}}(X_t)} c^T \lambda(T-t) \rho^T$$

is a matrix Brownian motion under $\mathbb{P}^{(\mathbf{K})}$.

We now write

$$\mathbb{E}[N_T] = \mathbb{E}[N_T \mathbb{1}_{\tau_{\mathbf{K}} > T}] + \mathbb{E}[N_T \mathbb{1}_{\tau_{\mathbf{K}} \leq T}].$$

By the dominated convergence theorem, we have $\mathbb{E}[N_T \mathbb{1}_{\tau_{\mathbf{K}} \leq T}] \rightarrow_{\mathbf{K} \rightarrow +\infty} 0$. Besides,

$$\mathbb{E}[N_T \mathbb{1}_{\tau_{\mathbf{K}} > T}] = \mathbb{E}[N_T^{(\mathbf{K})} \mathbb{1}_{\tau_{\mathbf{K}} > T}] = \mathbb{P}^{(\mathbf{K})}(\tau_{\mathbf{K}} > T),$$

and we have to prove that this probability goes to 1. To do so, we focus on the following SDE:

$$\begin{aligned} d\tilde{X}_t &= (\mathcal{Q} + (d-1)\epsilon^2 I_d^n + (b + 2\epsilon I_d^n g(T-t) + \epsilon I_d^n \rho \lambda^T (T-t) c) \tilde{X}_t \\ &\quad + \tilde{X}_t (b^T + 2\epsilon g(T-t) I_d^n + \epsilon c^T \lambda(T-t) \rho^T I_d^n)) dt \\ &\quad + \epsilon (\sqrt{\tilde{X}_t} dW_t I_d^n + I_d^n dW_t^T \sqrt{\tilde{X}_t}), \end{aligned}$$

starting from $\tilde{X}_0 = X_0$. We check that X solves before $\tau_{\mathbf{K}}$ and under $\mathbb{P}^{(\mathbf{K})}$ the same SDE as \tilde{X} under \mathbb{P} . This yields $\mathbb{P}^{(\mathbf{K})}(\tau_{\mathbf{K}} > T) = \mathbb{P}(\inf\{t \geq 0, \text{Tr}(\tilde{X}_t) \geq \mathbf{K}\} > T)$. Since the SDE satisfied by \tilde{X} is that of an affine diffusion on $\mathcal{S}_d^+(\mathbb{R})$, it is well defined for any $t \geq 0$. In particular, $\max_{t \in [0, T]} \text{Tr}(\tilde{X}_t) < \infty$ almost surely, which gives $\mathbb{P}(\inf\{t \geq 0, \text{Tr}(\tilde{X}_t) \geq \mathbf{K}\} > T) \rightarrow_{\mathbf{K} \rightarrow +\infty} 1$. \square

REMARK 3.3 The conditions (3.8) and (3.9) are satisfied for $\gamma = 0$ if and only if $-\Gamma \in \mathcal{S}_d^+(\mathbb{R})$ and, for all $t \geq 0$,

$$-\bar{\Gamma} - \frac{1}{2} c^T \lambda(t) \lambda^T(t) c \in \mathcal{S}_d^+(\mathbb{R}). \quad (3.12)$$

Since $|\lambda_i(t)| \leq \max(|\Lambda_i|, |\bar{\Lambda}_i/\kappa_i|)$, we obtain

$$\lambda(t)^T \lambda(t) \leq \sum_{i=1}^p \max(\Lambda_i^2, (\bar{\Lambda}_i/\kappa_i)^2).$$

We therefore have

$$\sum_{i=1}^p \max(\Lambda_i^2, (\bar{\Lambda}_i/\kappa_i)^2) I_d - \lambda(t)^T \lambda(t) \in \mathcal{S}_d^+(\mathbb{R}),$$

and then

$$\sum_{i=1}^p \max(\Lambda_i^2, (\bar{\Lambda}_i/\kappa_i)^2) c^T c - c^T \lambda(t) \lambda^T(t) c \in \mathcal{S}_d^+(\mathbb{R}).$$

Therefore, a sufficient condition for (3.12) is

$$-\bar{\Gamma} - \frac{1}{2} \sum_{i=1}^p \max(\Lambda_i^2, (\bar{\Lambda}_i/\kappa_i)^2) c^T c \in \mathcal{S}_d^+(\mathbb{R}).$$

With the Laplace transform (3.11), we have a mathematical tool to check if the process (X, Y) is stationary. This is important in terms of our modeling perspective: unless for some transitory period, one may expect that the factors are stable around some equilibrium. The next proposition gives a simple sufficient condition that ensures stationarity. It is proved in Appendix B (available online).

PROPOSITION 3.4 *If $-(b + b^T) \in \mathcal{S}_d^+(\mathbb{R})$ is positive definite, the process (X, Y) is stationary.*

REMARK 3.5 We chose to keep the dynamics of the process X in the space of positive semidefinite matrixes as general as possible. Choosing a Wishart specification for X (which corresponds to $\Omega = \epsilon^2 \alpha I_d^n$, $\alpha > 0$) does not lead to a significant simplification of the model. While Wishart processes admit an explicit Laplace transform, this is not the case for the process (X, Y) defined by (3.2). The drift term Ω allows us to account for the mean-reversion behavior of the process X ; we will typically consider a negative mean-reversion matrix b , in which case we can set $\Omega = -bx^\infty - x^\infty b^T$ so that the matrix process X mean reverts to x^∞ .

3.2 Model definition

DEFINITION 3.6 We assume that $(X_t, Y_t)_{t \geq 0}$ follows (3.1) and (3.2) under a risk-neutral measure. Then, we define the short interest rate by

$$r_t = \varphi + \sum_{i=1}^p Y_t^i + \text{Tr}(\gamma X_t), \quad (3.13)$$

with $\varphi \in \mathbb{R}$ and $\gamma \in \mathcal{S}_d(\mathbb{R})$.

From Proposition 3.2, we easily get the following result on the zero-coupon bonds.

COROLLARY 3.7 (Bond reconstruction formula) *Let $0 \leq t \leq T$ and let $P_{t,T} = \mathbb{E}[\exp(-\int_t^T r_s ds) \mid \mathcal{F}_t]$ denote the price at time t of a zero-coupon bond with maturity T . Let us assume that*

$$\gamma - \frac{1}{2} \left(\sum_{i=1}^p \frac{1}{\kappa_i^2} \right) c^T c \in \mathcal{S}_d^+(\mathbb{R}). \quad (3.14)$$

Then, using Remark 3.3, $P_{t,T}$ is given by

$$P_{t,T} = \exp(A(T-t) + \text{Tr}(D(T-t)X_t) + B(T-t)^T Y_t), \quad (3.15)$$

with $A(t) = \eta(t) - \varphi t$, $D(t) = g(t)$ and $B(t) = \lambda(t)$, where (η, g, λ) is the solution of (3.10) with (3.7), $\Lambda = 0$, $\Gamma = 0$, $\bar{\Gamma} = -\gamma$ and $\bar{\Lambda} = -\mathbf{1}_p$ (ie, $\bar{\Lambda}_i = -1$ for $1 \leq i \leq p$). In particular, we have $-D(T-t) \in \mathcal{S}_d^+(\mathbb{R})$.

Let us make some general comments on the model. It is close but slightly different to the model proposed in Bensussan (2010). Nonetheless, our presentation as a perturbation of the LGM gives us a better understanding of the model parameters. Thus, the vector process Y can be interpreted as in the LGM, meaning it is assumed to be the main driver of the yield curve. The individual factors are viewed as principal component movements of the yield curve. We chose to specify the model such that the matrix process X admits a similar interpretation. Typically, we will consider

$$\Omega = -bx^\infty - x^\infty b^T,$$

with b symmetric negative, to have a mean reversion toward a given covariance matrix x^∞ . The parameter ϵ measures the level of the perturbation around the LGM. The matrix process X plays the role of a stochastic variance–covariance matrix of the main movements of the yield curve. It is possible to define the diffusion parameter c such that the diagonal factors of the matrix X play the role of the instantaneous stochastic variance of the yield curve movement, and the off-diagonal terms play the role of the instantaneous covariance between two yield curve movements. The vector ρ is a correlation parameter between the processes Y and X . In a first approximation, interest rate options are options on linear combinations of the factors Y , and the instantaneous variance of these linear combinations amounts to linear combinations of the factors X .² Therefore, the correlation parameter ρ will drive the skew of interest rate options. We now make more precise comments on the model.

² Note that this is not completely true, even in the simple LGM. One important characteristic of the short-rate/factorial interest rate model is that the yield curve depends not only on the (stochastic) state variables of the model, but also on the volatility of the state variables. Therefore, the volatility factors X appear in the payoff of interest rate options.

- In order to keep the same factors as in the LGM, we would like to have $\gamma = 0$. However, this choice is possible only if the perturbation around the LGM is small enough, provided that $-(b + b^T)$ is positive definite (see Remark 3.8). Besides, even if $P_{t,T}$ is well defined for small enough $T - t$, it would then be given by the same formula; therefore, the yield curve dynamics depends on the factor X anyway.
- In order to have a clear interpretation of the volatility factor X on the factor Y , one possible choice is to consider $d = q \times p$ with $q \in \mathbb{N}^*$ and $c_{i,j} = \mathbb{1}_{(i-1) \times p < j \leq i \times p}$. Thus, from (3.4), the principal matrix $(X_{k,l})_{(i-1) \times p < k, l \leq i \times p}$ rules the instantaneous quadratic variation of the factor Y_i , while the submatrix $(X_{k,l})_{(i-1) \times p < k \leq i \times p, (j-1) \times p < l \leq j \times p}$ rules the instantaneous covariation between the factors Y^i and Y^j .
- The model does not prevent us from having a negative short rate or having $\mathbb{E}[|P_{t,T}|^k] = \infty$ for any $k > 0$, unless we consider the degenerated case ($p = 0$), where the yield curve is driven by the volatility factors X and the factors Y are null. This particular model has been studied by Gnoatto (2012).
- ATSMs generally consider constant parameters that are fixed over a large period and reflect the market behavior, while the current value of factors is fitted to market data. This is why we consider constant parameters here. However, in order to fit exactly zero-coupon bond prices, it is possible to take a time-dependent function φ while keeping the tractability of the model.

REMARK 3.8 Equation (3.14) is sufficient to ensure $P_{t,T}$ is well defined. However, this condition does not depend on ϵ ; we know that for $\epsilon = 0$, $P_{t,T}$ is well defined, since X is deterministic and Y is a Gaussian process. We can get a complementary sufficient condition when $-(b + b^T)$ is positive definite, which is a reasonable assumption, since it leads to a stationary process by Proposition 3.4. In this case, there exists $\mu > 0$ such that

$$-(b + b^T) - \mu I_d \in \mathcal{S}_d^+(\mathbb{R}).$$

By using Proposition 3.2 with $\Upsilon = (\mu/4\epsilon^2)I_d$, we get that (3.9) is satisfied if we have

$$\frac{\mu^2}{8\epsilon^2}I_d - \frac{\mu}{8\epsilon}(I_d^n\rho\lambda^Tc + I_d^n\rho\lambda^Tc) - \frac{1}{2}\lambda^T\lambda\lambda^Tc + \gamma \in \mathcal{S}_d^+(\mathbb{R})$$

for all $t \geq 0$. Since for $t \geq 0$, $\lambda(t)$ takes values in a compact subset of \mathbb{R}^p , there is $\epsilon_0 > 0$ such that this condition is satisfied for any $\epsilon \in (0, \epsilon_0)$.

REMARK 3.9 Let $a \in \mathcal{M}_d(\mathbb{R})$, and consider the model $r_t = \varphi + \sum_{i=1}^p Y_t^i + \text{Tr}(\tilde{\gamma} \tilde{X}_t)$, with

$$\begin{aligned} Y_t &= y + \int_0^t \kappa(\theta - Y_s) ds + \int_0^t \tilde{c} \sqrt{\tilde{X}_s} [\sqrt{1 - |\tilde{\rho}|^2} dZ_s + dW_s \tilde{\rho}], \\ \tilde{X}_t &= \tilde{x} + \int_0^t (\tilde{\Omega} + (d-1)\epsilon^2 a^T a + \tilde{b} \tilde{X}_s + \tilde{X}_s \tilde{b}^T) ds \\ &\quad + \epsilon \int_0^t \sqrt{\tilde{X}_s} dW_s a + a^T dW_s^T \sqrt{\tilde{X}_s}, \end{aligned}$$

and $\tilde{\gamma} \in \mathcal{S}_d(\mathbb{R})$, $\tilde{x}, \tilde{\Omega} \in \mathcal{S}_d^+(\mathbb{R})$, $\tilde{c}, \tilde{b} \in \mathcal{M}_d(\mathbb{R})$ and $\tilde{\rho} \in \mathbb{R}^d$ such that $|\tilde{\rho}| \leq 1$. This model may seem a priori more general, but this is not the case. In fact, let n be the rank of a , and let $u \in \mathcal{M}_d(\mathbb{R})$ be an invertible matrix such that $a^T a = (u^{-1})^T I_d^n (u^{-1})$. Then, $X_t = u^T \tilde{X}_t u$ solves

$$dX_t = [\Omega + (d-1)\epsilon^2 I_d^n + b X_t + X_t b^T] dt + u^T \sqrt{\tilde{X}_t} dW_t a u + u^T a^T dW_t^T \sqrt{\tilde{X}_t} u,$$

with $b = u^T \tilde{b} (u^{-1})^T$, $\Omega = u^T \tilde{\Omega} u \in \mathcal{S}_d^+(\mathbb{R})$ and starting from $x = u^T \tilde{x} u \in \mathcal{S}_d^+(\mathbb{R})$. After some calculations, we obtain

$$\langle d(Y_t)_m, d(Y_t)_{m'} \rangle = (\tilde{c} \tilde{X}_t \tilde{c}^T)_{m,m'} dt = (c X_t c^T)_{m,m'} dt$$

with $c = \tilde{c}(u^{-1})^T$;

$$\begin{aligned} \langle d(X_t)_{i,j}, d(X_t)_{k,l} \rangle &= \epsilon^2 [(X_t)_{i,k} \mathbb{1}_{j=l \leq n} + (X_t)_{i,l} \mathbb{1}_{j=k \leq n} \\ &\quad + (X_t)_{j,k} \mathbb{1}_{i=l \leq n} + (X_t)_{j,l} \mathbb{1}_{i=k \leq n}] dt; \end{aligned}$$

and

$$\begin{aligned} \langle d(Y_t)_m, d(X_t)_{i,j} \rangle &= \epsilon [(u^T \tilde{X}_t \tilde{c}^T)_{m,i} (u^T a^T \tilde{\rho})_j + (u^T \tilde{X}_t \tilde{c}^T)_{m,j} (u^T a^T \tilde{\rho})_i] dt \\ &= \epsilon [(X_t c^T)_{m,i} (u^T a^T \tilde{\rho})_j + (X_t c^T)_{m,j} (u^T a^T \tilde{\rho})_i] dt. \end{aligned}$$

Since the law of (X, Y) is characterized by its infinitesimal generator, we can assume without loss of generality that $\tilde{\rho} \in \ker(u^T a^T)^\perp = \text{Im}(au)$. Therefore, there is $\rho' \in \mathbb{R}^d$ such that $\tilde{\rho} = aup\rho'$, and we set $\rho_i = \rho'_i$ for $i \leq n$ and $\rho_i = 0$ for $n < i \leq d$. We have $|\rho|^2 = (\rho')^T I_d^n \rho' = |\tilde{\rho}|^2 \leq 1$; therefore, (X, Y) follows the same law as the solution of (3.1) and (3.2), and we have $r_t = \varphi + \sum_{i=1}^p Y_t^i + \text{Tr}(\gamma X_t)$ with $\gamma = u^{-1} \tilde{\gamma} (u^{-1})^T$.

3.3 Change of measure and Laplace transform

In the fixed income market, the pricing of vanilla products is often (if not always) made under a suitably chosen equivalent martingale measure that is different from

the risk-neutral measure. It is thus important to characterize the distribution of the underlying state variables under these measures. The forward-neutral measures are probably the most important example of such pricing measures. In what follows, we will see that the dynamics of the factors remains affine and keep the same structure under the forward measures.

3.3.1 Dynamics under the forward-neutral measures

We assume that (3.14) holds. Let Q^U denote the U -forward neutral probability, which is defined on \mathcal{F}_U by

$$\frac{dQ^U}{d\mathbb{P}} = \frac{\exp(-\int_0^U r_s ds)}{P_{0,U}}.$$

This is the measure associated with the numeraire $P_{t,U}$. It comes from the martingale property of discounted asset prices that for $t \in (0, U)$,

$$\begin{aligned} \frac{d(\exp(-\int_0^t r_s ds) P_{t,U})}{\exp(-\int_0^t r_s ds) P_{t,U}} &= 2\epsilon \text{Tr}(D(U-t)\sqrt{X_t} dW_t I_d^n) + B(U-t)^T c \sqrt{X_t} dW_t \rho \\ &\quad + \bar{\rho} B(U-t)^T c \sqrt{X_t} dZ_t \\ &= \text{Tr}([2\epsilon I_d^n D(U-t)\sqrt{X_t} + \rho B(U-t)^T c \sqrt{X_t}] dW_t) \\ &\quad + \bar{\rho} B(U-t)^T c \sqrt{X_t} dZ_t. \end{aligned}$$

From the Girsanov theorem, the processes

$$\begin{aligned} dW_t^U &= dW_t - \sqrt{X_t}(2\epsilon D(U-t)I_d^n + c^T B(U-t)\rho^T) dt, \\ dZ_t^U &= dZ_t - \bar{\rho} \sqrt{X_t} c^T B(U-t) dt \end{aligned}$$

are, respectively, matrix- and vector-valued Brownian motions under Q^U . They are independent. This yields the following dynamics for Y and X under Q^U :

$$\begin{aligned} dX_t &= (\Omega + (d-1)\epsilon^2 I_d^n + b^U(t)X_t + X_t(b^U(t))^T) dt \\ &\quad + \epsilon(\sqrt{X_t} dW_t^U I_d^n + I_d^n(dW_t^U)^T \sqrt{X_t}), \end{aligned} \tag{3.16}$$

$$\begin{aligned} dY_t &= \kappa(\theta - Y_t) dt + c X_t c^T B(U-t) dt + 2\epsilon c X_t D(U-t) I_d^n \rho dt \\ &\quad + c \sqrt{X_t} (dW_t^U \rho + \bar{\rho} dZ_t^U), \end{aligned} \tag{3.17}$$

with $b^U(t) = b + 2\epsilon^2 I_d^n D(U-t) + \epsilon I_d^n \rho B(U-t)^T c$.

3.3.2 Laplace transforms

We are now interested in calculating the law of (X_T, Y_T) under the U -forward measure for $T \leq U$. More precisely, we calculate $\mathbb{E}^{Q^U}[\exp(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T) | \mathcal{F}_t]$ for $t \in [0, T]$ by again using Proposition 3.2. We assume that (3.14) holds, and we have

$$\begin{aligned} & \mathbb{E}^{Q^U}[\exp(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T) | \mathcal{F}_t] \\ &= \frac{1}{P_{t,U}} \mathbb{E} \left[\exp(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T - (U-t)\varphi \right. \\ &\quad \left. - \int_t^U \mathbf{1}_p^T Y_s ds - \int_t^U \text{Tr}(\gamma X_s) ds \mid \mathcal{F}_t \right] \\ &= \frac{\mathbb{E}[\exp(\text{Tr}(\Gamma + D(U-T)X_T) + (\Lambda + B(U-T))^T Y_T \right. \\ &\quad \left. + A(U-T) - (T-t)\varphi - \int_t^T \mathbf{1}_p^T Y_s ds - \int_t^T \text{Tr}(\gamma X_s) ds) \mid \mathcal{F}_t]}{\exp(A(U-t) + \text{Tr}(D(U-t)X_t) + B(U-t)^T Y_t)}. \end{aligned}$$

We consider $\Gamma \in \mathcal{S}_d(\mathbb{R})$ and $\Lambda \in \mathbb{R}^p$, such that

$$-\Gamma \in \mathcal{S}_d^+(\mathbb{R}) \quad \text{and} \quad |\Lambda_i| \leq e^{-\kappa_i(U-T)}/\kappa_i \quad \text{for } 1 \leq i \leq p,$$

in order to have

$$|\Lambda_i + B_i(U-T)| \leq 1/\kappa_i$$

and

$$-(\Gamma + D(U-T)) \in \mathcal{S}_d^+(\mathbb{R}).$$

By Proposition 3.2, Equation (3.14) and Remark 3.3, we get that the expectation is finite and that

$$\begin{aligned} & \mathbb{E}^{Q^U}[\exp(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T) | \mathcal{F}_t] \\ &= \exp(A^U(t, T) + \text{Tr}(D^U(t, T)X_t) + B^U(t, T)^T Y_t), \quad (3.18) \end{aligned}$$

with $F^U(t, T) = \tilde{F}(T-t) + F(U-T) - F(U-t)$ for $F \in \{A, D, B\}$, where $(\tilde{B}, \tilde{D}, \tilde{A})$ is the solution of (3.10) with $\tilde{B}(0) = \Lambda + B(U-T)$, $\tilde{D}(0) = \Gamma + D(U-T)$, $\tilde{A}(0) = 0$, $\tilde{\Lambda} = \mathbf{1}_p$ and $\tilde{\Gamma} = -\gamma$.

COROLLARY 3.10 *Let (3.14) hold. For $\Gamma \in \mathcal{S}_d(\mathbb{R})$ and $\Lambda \in \mathbb{R}^p$ such that $-\Gamma \in \mathcal{S}_d^+(\mathbb{R})$ and $|\Lambda_i| \leq e^{-\kappa_i(U-T)}/\kappa_i$ for $1 \leq i \leq p$, $\mathbb{E}^{Q^U}[\exp(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T) | \mathcal{F}_t] < \infty$ almost surely for any $t \in [0, T]$ and is given by (3.18).*

Let us mention that, in practice, the formula above for $A^U(t, T)$, $D^U(t, T)$ and $B^U(t, T)$ requires us to solve two different ordinary differential equations (ODEs). It

may be more convenient to use the following one, which can be easily deduced from the dynamics of (X, Y) under the U -forward measure:

$$\begin{aligned}
 \frac{\partial B^U}{\partial t}(t, T) &= \kappa^T B^U(t, T), \quad B^U(T, T) = \Lambda, \\
 -\frac{\partial D^U}{\partial t}(t, T) &= 2\epsilon^2 D^U I_d^n D^U + D^U(b^U(t) + \epsilon I_d^n \rho(B^U)^T c) \\
 &\quad + (b^U(t) + \epsilon I_d^n \rho(B^U)^T c)^T D^U + \frac{1}{2} c^T B^U (B^U)^T c \\
 &\quad + c^T B^U B(U-t)^T c + \epsilon D(U-t) I_d^n \rho(B^U)^T c \\
 &\quad + \epsilon c^T B^U \rho^T I_d^n D(U-t), \quad D^U(T, T) = \Gamma, \\
 -\frac{\partial A^U}{\partial t}(t, T) &= B^U(t, T)^T \kappa \theta + \text{Tr}(D^U(t, T)(\Omega + \epsilon^2(d-1)I_d^n)), \\
 A^U(T, T) &= 0.
 \end{aligned} \tag{3.19}$$

4 EXPANSION OF THE VOLATILITY SMILE AROUND THE LINEAR GAUSSIAN MODEL

The goal of this section is to provide the asymptotic behavior of the caplet and swaption prices when the volatility parameter ϵ is close to zero. The practical interest of these formulas is to quickly give a proxy for these prices. Thus, they provide a tool for calibrating the model parameters to the smile. Let us mention here that Gram–Charlier-type expansions can also be applied to price caplets and swaptions, thanks to the affine structure of the model (see, for example, Collin-Dufresne and Goldstein 2002; Tanaka *et al* 2010). Some numerical examples are presented in Palidda (2015) for the pricing of caplets in our model. Here, we only present the expansion with respect to ϵ , since it is in accordance with our presentation of the model as a perturbation of the LGM.

The arguments that we use in this section to obtain the expansion have been developed in Fouque *et al* (2000). They rely on an expansion of the infinitesimal generator with respect to ϵ . Recently, this technique was applied by Bergomi and Guyon (2012) to provide an approximation under a multifactor model for the forward variance. Here, we have to take into account some specific features of the fixed income and work under the appropriate probability measure in order to apply these arguments. Unsurprisingly, the zero-order term in the expansion is exactly the volatility of the LGM with a time-dependent variance–covariance matrix. More interestingly, the higher-order terms allow us to confirm our intuition regarding the role of the parameters and factors that determine the shape and dynamics of the volatility.

Lastly, we should mention that the calculations presented in this section are rather formal. In particular, we implicitly assume that the caplet and swaption prices are smooth enough and admit expansions with respect to ϵ . A rigorous proof of these expansions is beyond the scope of this paper.

4.1 Price and volatility expansion for caplets

From (2.4), the only quantity of interest in order to understand the caplet volatility cube is what we call the forward caplet price,

$$\text{FCaplet}(t, T, \delta) = \mathbb{E}^{T+\delta}[(L_T(T, \delta) - K)^+ | \mathcal{F}_t],$$

which can be rewritten as a call option on the forward zero-coupon bond $P_{t,T}/P_{t,T+\delta}$:

$$\text{FCaplet}(t, T, \delta) = \frac{1}{\delta} \mathbb{E}^{T+\delta} \left[\left(\frac{P_{T,T}}{P_{T,T+\delta}} - (1 + \delta K) \right)^+ \middle| \mathcal{F}_t \right].$$

Since (X, Y) is a Markov process, $\text{FCaplet}(t, T, \delta)$ is a function of (X_t, Y_t) , and therefore we can define the forward price function:

$$P(t, x, y) = \mathbb{E}^{T+\delta} \left[\left(\frac{P_{T,T}}{P_{T,T+\delta}} - (1 + \delta K) \right)^+ \middle| X_t = x, Y_t = y \right]. \quad (4.1)$$

The goal of Section 4.1 is to obtain the second-order expansion (4.3) of P with respect to ϵ .

4.1.1 A convenient change of variable

We want to get an expansion of the caplet price with respect to ϵ . To do so, we need a priori to get an expansion to ϵ of the infinitesimal generator of the process (X, Y) under the probability $Q^{T+\delta}$. However, first we can make a change of variable that simplifies this approach. Thus, we define

$$H_t = \Delta A(t, T, \delta) + \text{Tr}(\Delta D(t, T, \delta) X_t) + \Delta B(t, T, \delta)^T Y_t,$$

with

$$\begin{aligned} \Delta A(t, T, \delta) &= A(t, T) - A(t, T + \delta), \\ (\Delta B, \Delta D)(t, T, \delta) &= (B, D)(T - t) - (B, D)(T + \delta - t). \end{aligned}$$

Thus, we have $P_{t,T}/(P_{t,T+\delta}) = e^{H_t}$. It is well known that $P_{t,T}/(P_{t,T+\delta})$ is a martingale under $Q^{T+\delta}$ (see, for example, Brigo and Mercurio 2006, Proposition 2.5.1). Thus, we get by Itô calculus from (3.16) and (3.17) that (X, H) solve the following SDEs:

$$\begin{aligned} dX_t &= (\Omega + \epsilon^2(d-1)I_d^n + b^{T+\delta}(t)X_t + X_t(b^{T+\delta}(t))^T)dt + \epsilon\sqrt{X_t}dW_t^{T+\delta}I_d^n \\ &\quad + \epsilon I_d^n(dW_t^{T+\delta})^T\sqrt{X_t}, \\ dH_t &= -\frac{1}{2}(\Delta B^T c X_t c^T \Delta B + 4\epsilon^2 \text{Tr}(\Delta D I_d^n \Delta D X_t) + 2\epsilon(\Delta B^T c X_t \Delta D I_d^n \rho))dt \\ &\quad + \Delta B^T c \sqrt{X_t}(dW_t^{T+\delta}\rho + \bar{\rho}dZ_t^{T+\delta}) + 2\epsilon \text{Tr}(\Delta D \sqrt{X_t}dW_t^{T+\delta}I_d^n). \end{aligned} \quad (4.2)$$

Therefore,

$$P(t, x, y) = \mathbb{E}^{T+\delta}[(e^{H_T} - (1 + \delta K))^+ | X_t = x, Y_t = y]$$

only depends on (x, y) through (x, h) , where

$$h = \Delta B(t, T, \delta)^T y + \text{Tr}(\Delta D(t, T, \delta)x) + \Delta A(t, T, \delta);$$

we still denote, by a slight abuse of notation, that

$$P(t, x, h) = \mathbb{E}^{T+\delta}[(e^{H_T} - (1 + \delta K))^+ | X_t = x, H_t = h].$$

Let us emphasize that this change of variable is crucial in order to apply an expansion procedure similar to that of Bergomi and Guyon (2012). It allows us to reduce the dimensionality of the underlying state variable. The variable H is one-dimensional, and it is the only variable that appears in the payoff of the caplet. Although this is obvious from the definition of the model, we insist that the implied volatility of caplets is a function of the factors X only. This can clearly be seen in the SDEs of (4.2); H_t can be viewed as a continuous version of the forward Libor, and its volatility depends on the factors X only.

4.1.2 Expansion of the price

From the SDEs (4.2), (3.4), (3.5) and (3.6), we get the following partial differential equation (PDE) representation of P :

$$\begin{aligned} \partial_t P + \mathcal{L}(t)P &= 0, \\ P(T, x, h) &= (e^h - (1 + \delta K))^+, \end{aligned}$$

where $\mathcal{L}(t)$ is the infinitesimal generator of (4.2). We assume that P admits a second-order expansion

$$P = P_0 + \epsilon P_1 + \epsilon^2 P_2 + o(\epsilon^2). \quad (4.3)$$

Our goal is to calculate, in a quite explicit way, the values of P_0 , P_1 and P_2 . We assume in our derivations that these functions are smooth enough. To determine the values of P_0 , P_1 and P_2 , we proceed as in Bergomi and Guyon (2012) and make an expansion of the generator $\mathcal{L}(t) = \mathcal{L}_0(t) + \epsilon \mathcal{L}_1(t) + \epsilon^2 \mathcal{L}_2(t) + \dots$ in order to obtain the PDEs satisfied by P_0 , P_1 and P_2 . Namely, we obtain

$$\begin{aligned} \partial_t P_0 + \mathcal{L}_0(t)P_0 &= 0, & P_0(T, x, h) &= (e^h - (1 + \delta K))^+, \\ \partial_t P_1 + \mathcal{L}_0(t)P_1 + \mathcal{L}_1(t)P_0 &= 0, & P_1(T, x, h) &= 0, \\ \partial_t P_2 + \mathcal{L}_0(t)P_2 + \mathcal{L}_1(t)P_0 + \mathcal{L}_2(t)P_1 &= 0, & P_2(T, x, h) &= 0. \end{aligned}$$

Thus, we can first solve the PDE for P_0 , then for P_1 and so on. Let $\text{BS}(h, v) = \mathbb{E}[(\exp(h - \frac{1}{2}v + \sqrt{v}G) - (1 + \delta K))^+]$, with $G \sim N(0, 1)$, denote the Black–Scholes price with realized volatility v . We obtain easily that

$$P_0(t, x, h) = \text{BS}(h, v(t, T, \delta, x)),$$

with

$$v(t, T, \delta, x) = \int_t^T \Delta B(u, T, \delta)^T c X_{u-t}^0(x) c^T \Delta B(u, T, \delta) du, \quad (4.4)$$

$$X_s^0(x) = e^{bs} \left(x + \int_0^s e^{-bu} \Omega e^{-b^T u} du \right) e^{b^T s}. \quad (4.5)$$

The higher-order terms are given by³

$$P_1(t, x, h) = (c_1(t, T, \delta, x)(\partial_h^3 - \partial_h^2) + c_2(t, T, \delta, x)(\partial_h^2 - \partial_h)) P_0(t, x, h) \quad (4.6)$$

and

$$\begin{aligned} P_2(t, x, h) = & [(d_1(t, T, \delta, x)(\partial_h^2 - \partial_h)^2 + d_2(t, T, \delta, x)(\partial_h^2 - \partial_h)\partial_h \\ & + d_3(t, T, \delta, x)(\partial_h^2 - \partial_h)) + (e_1(t, T, \delta, x)(\partial_h^2 - \partial_h)^2\partial_h^2 \\ & + e_2(t, T, \delta, x)(\partial_h^2 - \partial_h)^2\partial_h + e_3(t, T, \delta, x)(\partial_h^2 - \partial_h)^2 \\ & + e_4(t, T, \delta, x)(\partial_h^2 - \partial_h)\partial_h^2 + e_5(t, T, \delta, x)(\partial_h^2 - \partial_h)\partial_h \\ & + e_6(t, T, \delta, x)(\partial_h^2 - \partial_h))] P_0(t, x, h). \end{aligned} \quad (4.7)$$

The coefficients c_i , d_i and e_i are given in Appendix A.1 (available online). We recall that the derivatives $\partial_h^i P_0$ of P_0 with respect to h can be calculated explicitly, so the expansion is very efficient from the point of view of the computational time (see Section 6).

REMARK 4.1 It is easy to obtain, then, the expansion $v_{\text{Imp}} = v_0 + \epsilon v_1 + \epsilon^2 v_2 + o(\epsilon^2)$ of the implied volatility defined by $\delta \text{FCaplet}(t, T, \delta) = \text{BS}(h, v_{\text{Imp}})$. We obtain, as expected, $v_0 = v(t, T, \delta, x)$ and

$$\frac{v_1}{2} = c_2(t, T, \delta, x) + c_1(t, T, \delta, x) \left(\frac{1}{2} - \frac{h - \log(1 + \delta K)}{v_0} \right). \quad (4.8)$$

Since neither c_1 nor c_2 depends on the strike, the skew is at the first order in ϵ proportional to c_1 , which is, in turn, a linear function of ρ . In particular, we have a flat smile at the first order when $\rho = 0$, as one may expect.

³ The details of these simple but tedious calculations for the caplets and swaptions are available online at <http://bit.ly/2grtw71>, in the first draft of this paper.

4.2 Price and volatility expansion for swaptions

From (2.5), the only quantity of interest with regard to understanding the swaption volatility cube is what we call the annuity-forward swaption price:

$$\text{AFSwaption}(t, T, m, \delta) = \mathbb{E}^A[(S_t(T, m, \delta) - K)^+ | \mathcal{F}_t].$$

It is standard to view swaptions as a basket option of forward Libors with stochastic weights. We have

$$S_t(T, m, \delta) = \sum_{i=1}^m \omega_t^i L_t(T + (i-1)\delta, T + i\delta), \quad (4.9)$$

$$\omega_t^i = \frac{P_{t, T+i\delta}}{\sum_{i=1}^m P_{t, T+i\delta}}. \quad (4.10)$$

The difficulty here comes from the fact that forward Libors and stochastic weights are complicated functions of the state variables (X, Y) . The first implication of this is that the change of measure between \mathbb{P} and \mathbb{Q}^A is also complicated, and the dynamics of the state variables under this new measure are quite unpleasant to work with. The second implication is that we cannot directly operate a convenient change of variable as we did for caplets. Thus, in order to derive an expansion for swaptions, we proceed stepwise. First, we use a standard approximation that freezes the weights at their initial value (see, for example, Brigo and Mercurio 2006, p. 239; D'Aspremont 2003; Piterbarg 2009). This is justified by the fact that the variation of the weights is less important than the variation of the forward Libors.⁴ Second, we use a similar approximation for the swap rate. Thus, the approximated swap rate is an affine function of the underlying state variables, which enables us to take advantage of the affine structure of the model. Let us mention that this technique is similar to the quadratic approximation of the swap rate proposed by Piterbarg (2009). Finally, we perform our expansion on the affine approximation of the swap rates and obtain the second-order expansion (4.17), which is the main result of Section 4.2.

4.2.1 Dynamics of the factors under the annuity measure

The annuity measure knowing the information up to date t , $\mathbb{Q}^A | \mathcal{F}_t$, is defined by

$$\frac{d\mathbb{Q}^A}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \exp \left(- \int_t^T r_s ds \right) \frac{A_T(T, m, \delta)}{A_t(T, m, \delta)}.$$

⁴ To the best of our knowledge, there have been very few attempts to quantify this statement either theoretically or numerically. In D'Aspremont (2003), the author investigates the accuracy of the approximation for pricing swaptions in the lognormal Brace–Gatarek–Musiela (BGM) model. He shows that the approximation is less efficient for long maturities and long tenors.

This comes from the martingale property of discounted asset prices under the risk-neutral measure that

$$\frac{d(\exp(-\int_0^t r_s ds) A_t(T, m, \delta))}{\exp(-\int_0^t r_s ds) A_t(T, m, \delta)} = \sum_{i=1}^m \omega_t^i (B(T + i\delta - t)^T c \sqrt{X_t} (dW_t \rho + \bar{\rho} dZ_t) + 2\epsilon \text{Tr}(D(T + i\delta - t) \sqrt{X_t} dW_t I_d^n)). \quad (4.11)$$

From the Girsanov theorem, the change of measure is given by

$$\begin{aligned} dW_t^A &= dW_t - \sqrt{X_t} \left(2\epsilon \sum_{i=1}^m \omega_t^i D(T + i\delta - t) I_d^n + c^T B(T + i\delta - t) \rho^T \right) dt, \\ dZ_t^A &= dZ_t - \bar{\rho} \sqrt{X_t} c^T \sum_{i=1}^m \omega_t^i B(T + i\delta - t) dt. \end{aligned}$$

This allows us to calculate from (3.1) and (3.2) the dynamics of the state variables under the annuity measure \mathcal{Q}^A :

$$\begin{aligned} dY_t &= \left(\kappa(\theta - Y_t) + c X_t c^T \sum_{k=1}^m \omega_t^k B(T + k\delta - t) \right. \\ &\quad \left. + 2\epsilon c X_t \sum_{k=1}^m \omega_t^k D(T + k\delta - t) I_d^n \rho \right) dt \\ &\quad + c \sqrt{X_t} (\bar{\rho} dZ_t^A + dW_t^A \rho), \end{aligned} \quad (4.12)$$

$$\begin{aligned} dX_t &= (\Omega + \epsilon^2 (d-1) I_d^n + b^A(t) X_t + X_t (b^A(t))^T) dt \\ &\quad + \epsilon (\sqrt{X_t} dW_t^A I_d^n + I_d^n (dW_t^A)^T \sqrt{X_t}), \end{aligned} \quad (4.13)$$

where

$$b^A(t) = b + \epsilon I_d^n \rho \sum_{k=1}^m \omega_t^k B(T + k\delta - t)^T c + 2\epsilon^2 I_d^n \sum_{k=1}^m \omega_t^k D(T + k\delta - t).$$

4.2.2 An affine approximation of the forward swap rate

The forward swap rate is a martingale under the annuity measure \mathcal{Q}^A . Therefore, we can only focus on the martingale terms when applying Itô's formula to

$$\frac{P_{t,T} - P_{t,T+m\delta}}{\sum_{i=1}^m P_{t,T+i\delta}};$$

we get from (3.15) that

$$\begin{aligned}
 dS_t(T, m, \delta) &= \left[\omega_t^0 B(T-t)^T - \omega_t^m B(T+m\delta-t)^T - S_t(T, m, \delta) \sum_{k=1}^m \omega_t^k B(T+k\delta-t)^T \right] \\
 &\quad \times c \sqrt{X_t} (dW_t^A \rho + \bar{\rho} dZ_t^A) \\
 &\quad + 2\epsilon \text{Tr} \left(\left[\omega_t^0 D(T-t) - \omega_t^m D(T+m\delta-t) \right. \right. \\
 &\quad \left. \left. - S_t(T, m, \delta) \sum_{k=1}^m \omega_t^k D(T+k\delta-t) \right] \sqrt{X_t} dW_t^A I_d^n \right). \tag{4.14}
 \end{aligned}$$

Using a slight abuse of notation, we will now drop the (T, m, δ) dependence of the swap rate and simply denote by S_t its time t value. We now use the standard approximation that consists of freezing the weights ω_t^k and the value of the swap rate S_t on the right-hand side to their respective values at zero. We then have

$$dS_t = B^S(t)^T c \sqrt{X_t} (dW_t^A \rho + \bar{\rho} dZ_t^A) + 2\epsilon \text{Tr}(D^S(t) \sqrt{X_t} dW_t^A I_d^n), \tag{4.15}$$

where

$$\begin{aligned}
 (B, D)^S(t) &= \omega_0^0 (B, D)(T-t) - \omega_0^m (B, D)(T+m\delta-t) \\
 &\quad - S_0(T, m, \delta) \sum_{k=1}^m \omega_0^k (B, D)(T+k\delta-t).
 \end{aligned}$$

These coefficients are time dependent and deterministic. We do the same approximation on X and get

$$\begin{aligned}
 dX_t &= (\Omega + \epsilon^2(d-1)I_d^n + b_0^A(t)X_t + X_t(b_0^A(t))^T) dt \\
 &\quad + \epsilon(\sqrt{X_t} dW_t^A I_d^n + I_d^n (dW_t^A)^T \sqrt{X_t}),
 \end{aligned} \tag{4.16}$$

where

$$b_0^A(t) = b + \epsilon I_d^n \rho \sum_{k=1}^m \omega_0^k B(T+k\delta-t)^T c + 2\epsilon^2 I_d^n \sum_{k=1}^m \omega_0^k D(T+k\delta-t).$$

Thanks to this approximation, we remark that the process, which we still denote by (S_t, X_t) for simplicity, is now affine. This enables us to use the same argument as for the caplet prices in order to get an expansion of the price. The only difference lies in the fact that the expansion is around the Gaussian rather than the lognormal model.

4.2.3 The swaption price expansion

Let $P^S(t, x, s) = \mathbb{E}^A[(S_t - K)^+ | S_t = s, X_t = x]$ denote the price of the swaption at time $t \in [0, T]$. This solves the following pricing PDE:

$$\partial_t P^S + \mathcal{L}(t)P^S = 0, \quad t \in (0, T), \quad P^S(T, x, s) = (s - K)^+,$$

where \mathcal{L} is the infinitesimal generator of the SDEs (4.15) and (4.16). Again, we assume that P^S admits a second-order expansion

$$P^S = P_0^S + \epsilon P_1^S + \epsilon^2 P_2^S + o(\epsilon^2), \quad (4.17)$$

and that the functions P_0^S , P_1^S and P_2^S are smooth enough. Let $\text{BH}(s, v) = \mathbb{E}[(s + \sqrt{v}G - K)^+]$ with $G \sim N(0, 1)$ denote the European call price with strike K in Bachelier's model, with realized volatility $v > 0$ and spot price $s \in \mathbb{R}$. We obtain

$$P_0^S(t, x, s) = \text{BH}(s, v^S(t, T, x)), \quad (4.18)$$

where

$$v^S(t, T, x) = \int_t^T B^S(r)^T c X_{r-t}^0(x) c^T B^S(r) dr \quad (4.19)$$

and $X_s^0(x)$ is defined by (4.5). The higher-order terms are

$$P_1^S(t, x, s) = (c_1^S(t, T, x)\partial_s^3 + c_2^S(t, T, x)\partial_s^2)\text{BH}(s, v^S(t, T, x)), \quad (4.20)$$

$$\begin{aligned} P_2^S(t, x, s) = & [d_1^S(t, T, x)\partial_s^4 + d_2^S(t, T, x)\partial_s^3 + d_3^S(t, T, x)\partial_s^2 \\ & + e_1^S(t, T, x)\partial_s^6 + e_2^S(t, T, x)\partial_s^5 + e_3^S(t, T, x)\partial_s^4 \\ & + e_4^S(t, T, x)\partial_s^4 + e_5^S(t, T, x)\partial_s^3 + e_6^S(t, T, x)\partial_s^2] \\ & \times \text{BH}(s, v^S(t, T, x)), \end{aligned} \quad (4.21)$$

where the coefficients c_i^S , d_i^S and e_i^S are given in Appendix A.2 (available online). Again, the derivatives of P^S with respect to s can be calculated explicitly, which makes this formula very efficient from a computational point of view.

4.3 Numerical results

We now assess the accuracy of the expansions we have developed in some examples. In practice, we are interested in knowing up to what level of parameters and for what sets of maturities and tenors the accuracy of the expansion is satisfactory. Let us recall that our expansion for caplets results from the combination of two expansions: the first on the support matrix function D up to order 1 in ϵ is given by (52) and (53) in Appendix A.1 (available online), while the second on the infinitesimal generator of the Markov

process (X, H) is defined by (4.2). By construction, the approximation of $D(\tau)$ will be more accurate for a small τ . As a consequence, for a given set of parameters, the full expansion will likely be more accurate for short-maturity, short-tenor caplets. The expansion for swaptions results from a supplementary approximation step, which consists of freezing the weights ω^i in the diffusion of the Markov process (X, S) defined by (4.13) and (4.14). This approximation can be inaccurate for long-maturity and long-tenor swaptions. Therefore, we expect the full expansion to be more accurate for short-maturity, short-tenor swaptions.

We assess the quality of the price expansion for caplets and swaptions. We compare the expanded price with the price computed using Monte Carlo simulation and discretization Scheme 1, described in Section 5, on a regular time grid. The expanded prices and the Monte Carlo prices are compared in terms of the normal implied volatility of the forward Libor for caplets and the forward swap rate for swaptions. The implied volatility is given in basis points (bps) (10^{-4}). In abscissa is indicated the difference between the strike and the at-the-money value, and the unit is one percent. A $6M \times 2Y$ caplet will denote a caplet with maturity $T = 2$ years and tenor $\delta = 0.5$ years, while a $5Y \times 2Y$ swaption will denote a swaption with maturity $T = 2$ years and tenor $m\delta = 5$ years.

We have tested different sets of model parameters. The parameters values have been chosen in such a way that the yield curve and volatility levels generated by the model are in line with today's USD and EUR interest rate market levels. Here, we only consider the following parameter set, with $p = 2$ and $d = 2$:

$$\begin{aligned}\kappa &= \text{diag}(0.1, 1), \quad c = I_d, \quad b = -\text{diag}(0.41, 0.011), \\ \Omega &= -(bx_\infty + x_\infty b^T) + 0.4I_d, \quad \gamma = 0.001I_d, \\ x &= 10^{-4} \begin{pmatrix} 2.25 & -1.2 \\ -1.2 & 1 \end{pmatrix}, \quad x_\infty = 10^{-4} \begin{pmatrix} 1 & -0.125 \\ -0.125 & 0.25 \end{pmatrix}. \end{aligned} \quad (4.22)$$

We note that $-(b + b^T) = -2b$ is positive definite. We know from Remark 3.8 that the condition of non-explosion will be verified in general for this set of parameters when ϵ is small enough, and we have checked that the yield curve given by this set of parameters is well defined up to fifty years.

In all of the figures, the dotted line gives the Monte Carlo smile obtained with 100 000 simulation paths, the solid line with small arrows is the expanded smile and the two continuous solid lines are the upper and lower bounds of the 95% confidence interval of the Monte Carlo price. Figures 2 and 3 show the accuracy of the expansion for the valuation of caplets. The approximation is accurate for expiries up to two years, and less accurate with the same parameters for longer expiries. For maturities up to two years, the at-the-money volatility of the expanded smile is almost identical to the Monte Carlo smile, and the whole expanded smile stays within the 95% confidence

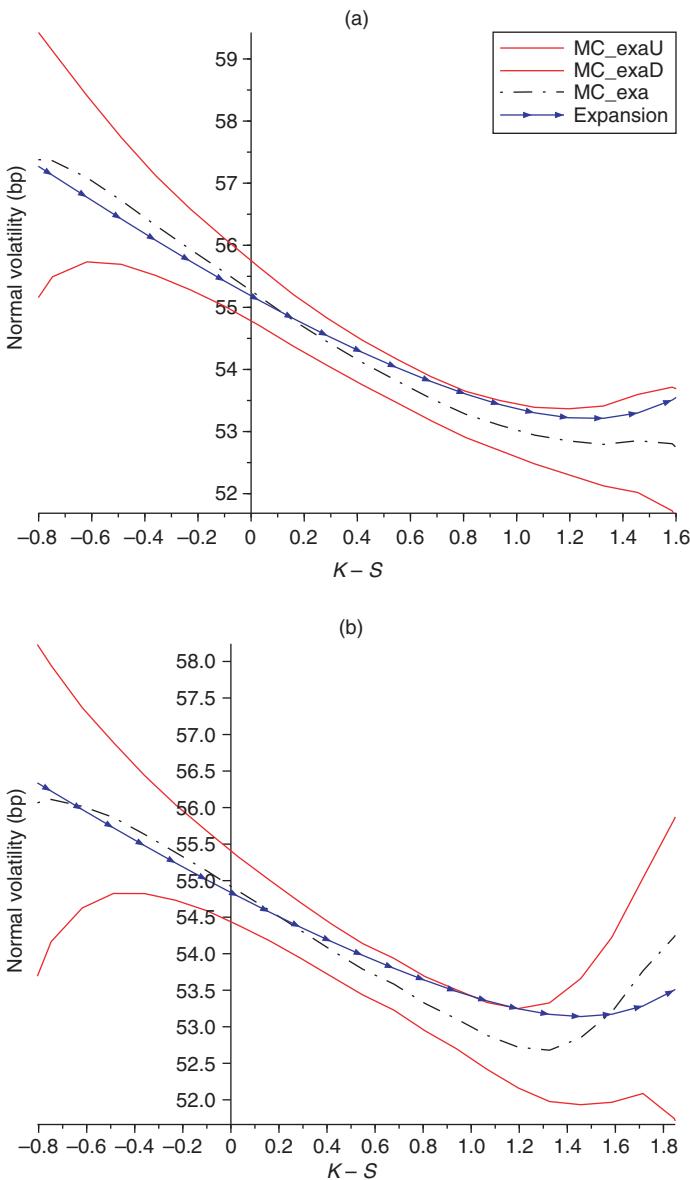
interval. Figure 4 shows the accuracy of the expansion for the valuation of swaptions. We observe that the expansion is more accurate for negative values of the correlation parameters ρ (a similar behavior is observed for caplets). This can be intuitively understood from the Riccati equation (3.10): a negative ρ pushes D to zero, while a positive one pushes D away from zero. The expansion that we use on D (see (52) and (53) in Appendix A.1, available online) is then less accurate. Overall, the expansion is accurate at-the-money and is much less accurate out-of-the-money. For example, Figure 3(b) shows that the expanded smile of the six-month maturity/five-year expiry smile is quite inaccurate, and the expanded smile fails to fit the skew of the Monte Carlo smile. However, the difference in at-the-money volatility between the expanded price and the Monte Carlo is around 1bp.

To sum up, the second-order expansion is basically accurate for small perturbations and small maturities. Otherwise, one should be careful and rely on other methods such as the Monte Carlo method or Fourier inversion method. Nonetheless, as discussed in Section 6, the calculation of the expansion is much faster than in the other methods. Thus, it may be relevant to start a calibration routine and select a reasonable set of parameters.

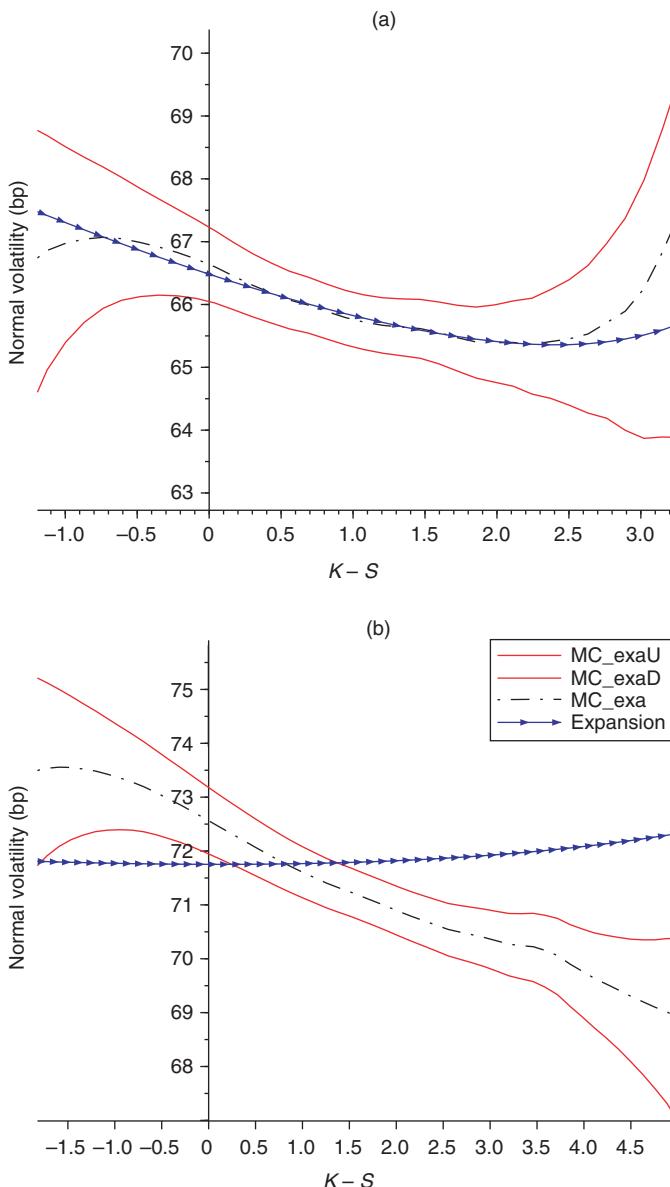
5 SECOND-ORDER DISCRETIZATION SCHEMES FOR MONTE CARLO SIMULATION

The goal of this section is to construct discretization schemes for the process (X, Y) , defined by (3.1) and (3.2). It is crucial to have an efficient way to simulate the model in order to use it in practice. Ideally, the model should be calibrated to market data for vanilla options, such as caplets and swaptions, and then be used to calculate exotic option prices. The calculation of these prices is generally made with a Monte Carlo algorithm, which requires us to simulate the process (X, Y) .

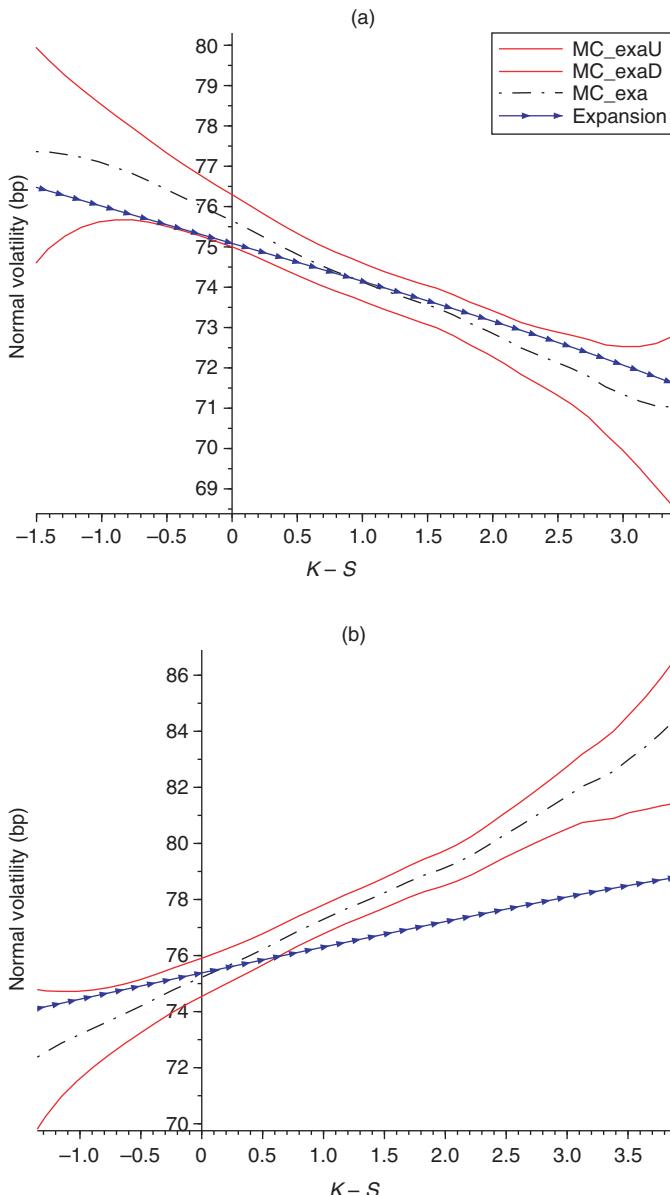
It is worth recalling that the standard Euler–Maruyama scheme is not well defined for square-root diffusions, even in dimension one (see Alfonsi (2010)). Thus, we have to consider a different scheme. Here, we use the splitting technique that has already been used by Ahdida and Alfonsi (2013) for Wishart processes. We explain briefly the main line of this method and refer to Alfonsi (2010) for precise statements in a framework that embeds affine diffusions. Let us consider that we want to approximate an SDE ξ with infinitesimal generator \mathcal{L} on the regular time grid $t_i = i T/N$ for $i = 0, \dots, N$. A scheme is fully described by a probability law $\hat{p}_x(t, dz)$ that approximates the law of ξ_t given $\xi_0 = x$. We denote by $\hat{\xi}_t^x$ a random variable following this law. Then, the law of the corresponding discretization scheme $(\hat{\xi}_{t_i}, 0 \leq i \leq N)$ is given by $\hat{\xi}_0 = \xi_0$ and $\hat{p}_{\hat{\xi}_{t_i}}(T/N, dz)$ is the conditional law of $\hat{\xi}_{t_{i+1}}$ given $(\hat{\xi}_{t_j}, 0 \leq j \leq i)$. Then, we would like to know the error made when using the approximation scheme instead of the original process ξ . We basically have the following result, up to technical

FIGURE 2 Smile of a 1Y × 1Y caplet.

$\rho^T = (-0.4, -0.2)$. Plot of the smile obtained with the expansion against the Monte Carlo smile obtained with 100 000 paths and a discretization grid of four points for different values of the parameter ϵ : (a) $\epsilon = 0.002$; (b) $\epsilon = 0.0015$. The forward Libor value is $L(0, 1Y, 1Y) = 1.02\%$.

FIGURE 3 Smile of a $6M \times 2Y$ caplet and of a $6M \times 5Y$ caplet.

$\rho^T = (-0.4, -0.2)$ and $\epsilon = 0.0015$. (a) Plot of the smile of a $6M \times 2Y$ caplet against the Monte Carlo smile. (b) Plot of the expanded smile of a $6M \times 5Y$ caplet against the Monte Carlo smile. The Monte Carlo smile is obtained with 100 000 paths and a discretization grid of eight points. The forward Libor values are $L(0, 6M, 2Y) = 1.14\%$ and $L(0, 6M, 5Y) = 1.35\%$.

FIGURE 4 Smile of a 5Y × 2Y swaption.

$\epsilon = 0.0015$. Plot of the expanded smile of a 5Y × 2Y swaption with a coupon payment frequency of six months against the Monte Carlo smile obtained with 100 000 paths and a discretization grid of eight points for different values of the parameter ρ : (a) $\rho^T = (-0.4, -0.2)$; (b) $\rho^T = (0.4, 0.2)$. The forward swap rate value is $S(0, 5Y, 2Y) = 1.3\%$.

details that are given in Alfonsi (2010). If $\hat{\xi}_t^x$ satisfies the expansion

$$\mathbb{E}[f(\hat{\xi}_t^x)] = f(x) + t\mathcal{L}f(x) + \frac{1}{2}t^2\mathcal{L}^2f(x) + O(t^3) \quad (5.1)$$

for any smooth function f , then

$$\exists C > 0, \quad |\mathbb{E}[f(\hat{\xi}_{tN})] - \mathbb{E}[f(\hat{\xi}_{tN})]| \leq C/N^2.$$

Thus, to get a weak error of order 2, we mainly have to construct a scheme $\hat{\xi}_t^x$ that satisfies (5.1). We can iteratively construct second-order schemes by splitting the infinitesimal generator. In fact, let us assume that $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$ and that $\hat{\xi}_t^{i,x}$ is a second-order scheme for \mathcal{L}_i . Let B be an independent Bernoulli variable with parameter 1/2. Then, the schemes

$$\hat{\xi}_{t/2}^{1,\hat{\xi}_t^{2,\hat{\xi}_t^{1,x}}/2} \quad \text{and} \quad B\hat{\xi}_t^{2,\hat{\xi}_t^{1,x}} + (1-B)\hat{\xi}_t^{1,\hat{\xi}_t^{2,x}} \quad (5.2)$$

satisfy (5.1) and are thus second-order schemes for \mathcal{L} . Therefore, one strategy for constructing a second-order scheme is to split the infinitesimal generator into elementary pieces for which second-order schemes or even exact schemes are known.

To use this splitting technique, we first have to calculate the infinitesimal generator of (X, Y) . It is defined for a C^2 function $f: \mathcal{M}_d(\mathbb{R}) \times \mathbb{R}^p \rightarrow \mathbb{R}$ by $\mathcal{L}f(x, y) = \lim_{t \rightarrow 0+} ((\mathbb{E}[f(X_t, Y_t)] - f(x, y))/t)$. From (3.4), (3.5) and (3.6), we easily get

$$\begin{aligned} \mathcal{L} = & \sum_{m=1}^p (\kappa(\theta - y))_m \partial_{y_m} + \sum_{1 \leq i, j \leq d} (\Omega + (d-1)\epsilon^2 I_d^n + bx + xb^T)_{i,j} \partial_{x_{i,j}} \\ & + \frac{1}{2} \sum_{m, m'=1}^p (c x c^T)_{m, m'} \partial_{y_m} \partial_{y_{m'}} \\ & + \frac{1}{2} \sum_{m=1}^p \sum_{1 \leq i, j \leq d} \epsilon [(cx)_{m,i} (I_d^n \rho)_j + (cx)_{m,j} (I_d^n \rho)_i] \partial_{x_{i,j}} \partial_{y_m} \\ & + \frac{1}{2} \sum_{1 \leq i, j, k, l \leq d} \epsilon^2 [x_{i,k} (I_d^n)_{j,l} + x_{i,l} (I_d^n)_{j,k} \\ & \quad + x_{j,k} (I_d^n)_{i,l} + x_{j,l} (I_d^n)_{i,k}] \partial_{x_{i,j}} \partial_{x_{k,l}}. \end{aligned}$$

Here, ∂_{y_m} denotes the partial derivative with respect to the m th coordinate in \mathbb{R}^p , and $\partial_{x_{i,j}}$ denotes the partial derivative with respect to the element at the i th row and j th column. When $\rho = 0$, this operator is simply the sum of the infinitesimal generators for X and the generator for Y when X is frozen. We know a second-order scheme for X from Ahdida and Alfonsi (2013). When X is frozen, Y follows

an Ornstein–Uhlenbeck process, and the law of Y_t is a Gaussian vector that can be sampled exactly. By using the composition rule (5.2), we get a second-order scheme for (X, Y) .

Thus, the difficulty here comes from the correlation between X and Y , which has to be handled with care. First, we make some simplifications. The first term

$$\sum_{m=1}^p (\kappa(\theta - y))_m \partial_{y_m}$$

is the generator of the linear ODE $y'(t) = \kappa(\theta - y(t))$, which is solved exactly by $y(t) = e^{-\kappa t} y(0) + (I_p - e^{-\kappa t})\theta$. Therefore, it is sufficient to have a second-order scheme for $\mathcal{L} - \sum_{m=1}^p (\kappa(\theta - y))_m \partial_{y_m}$, which is the generator of (3.1) and (3.2) when $\kappa = 0$. When $\kappa = 0$, we have $Y_t = y + c(\tilde{Y}_t - \tilde{Y}_0)$, with

$$\tilde{Y}_t = \tilde{Y}_0 + \int_0^t \sqrt{X_s} [\bar{\rho} dZ_s + dW_s \rho].$$

We can then focus on getting a second-order scheme for (X, \tilde{Y}) , which amounts to working with $p = d$ and $c = I_d$. It is therefore sufficient to find a second-order scheme for the SDE:

$$\begin{aligned} Y_t &= y + \int_0^t \sqrt{X_s} [\bar{\rho} dZ_s + dW_s \rho], \\ X_t &= x + \int_0^t (\Omega + (d-1)\epsilon^2 I_d^n + bX_s + X_s b^T) ds \\ &\quad + \epsilon \int_0^t \sqrt{X_s} dW_s I_d^n + I_d^n dW_s^T \sqrt{X_s}, \end{aligned}$$

with the infinitesimal generator

$$\begin{aligned} \mathcal{L} = & \sum_{1 \leq i, j \leq d} (\Omega + (d-1)\epsilon^2 I_d^n + bx + xb^T)_{i,j} \partial_{x_{i,j}} \\ & + \frac{1}{2} \sum_{m=1}^d \sum_{1 \leq i, j \leq d} \epsilon [x_{m,i} (I_d^n \rho)_j + x_{m,j} (I_d^n \rho)_i] \partial_{x_{i,j}} \partial_{y_m} \\ & + \frac{1}{2} \sum_{m,m'=1}^d x_{m,m'} \partial_{y_m} \partial_{y_{m'}} \\ & + \frac{1}{2} \sum_{1 \leq i, j, k, l \leq d} \epsilon^2 [x_{i,k} (I_d^n)_{j,l} + x_{i,l} (I_d^n)_{j,k} \\ & \quad + x_{j,k} (I_d^n)_{i,l} + x_{j,l} (I_d^n)_{i,k}] \partial_{x_{i,j}} \partial_{x_{k,l}}. \end{aligned} \quad (5.3)$$

5.1 A second-order scheme

For $1 \leq q \leq d$, we define $e_d^q \in \mathcal{S}_d^+(\mathbb{R})$ by $(e_d^q)_{k,l} = \mathbb{1}_{k=l=q}$ and $g_d^q \in \mathbb{R}^d$ by $(g_d^q)_k = \mathbb{1}_{q=k}$, so that

$$I_d^n = \sum_{q=1}^n e_d^q \quad \text{and} \quad I_d^n \rho = \sum_{q=1}^n \rho_q g_d^q.$$

We define

$$\begin{aligned} \mathcal{L}_q^c &= \epsilon^2(d-1)\partial_{x_{q,q}} + \frac{1}{2} \sum_{m=1}^d \sum_{1 \leq i,j \leq d} \epsilon \rho_q [x_{m,i}(g_d^q)_j + x_{m,j}(g_d^q)_i] \partial_{x_{i,j}} \partial_{y_m} \\ &\quad + \frac{\rho_q^2}{2} \sum_{m,m'=1}^d x_{m,m'} \partial_{y_m} \partial_{y_{m'}} \\ &\quad + \frac{1}{2} \sum_{1 \leq i,j,k,l \leq d} \epsilon^2 [x_{i,k}(e_d^q)_{j,l} + x_{i,l}(e_d^q)_{j,k} \\ &\quad \quad \quad + x_{j,k}(e_d^q)_{i,l} + x_{j,l}(e_d^q)_{i,k}] \partial_{x_{i,j}} \partial_{x_{k,l}}. \end{aligned} \tag{5.4}$$

We consider the splitting $\mathcal{L} = \mathcal{L}' + \mathcal{L}'' + \sum_{q=1}^n \mathcal{L}_q^c$ of the operator (5.3), with

$$\begin{aligned} \mathcal{L}' &= \sum_{1 \leq i,j \leq d} (\Omega + bx + xb^T)_{i,j} \partial_{x_{i,j}}, \\ \mathcal{L}'' &= \left(1 - \sum_{q=1}^n \rho_q^2\right) \frac{1}{2} \sum_{m,m'=1}^d x_{m,m'} \partial_{y_m} \partial_{y_{m'}}. \end{aligned}$$

The operator \mathcal{L}' is that of the linear ODE

$$x'(t) = \Omega + (d-1)\epsilon^2 I_d^n + bx + xb^T,$$

which can be solved exactly and stays in the set of semidefinite positive matrixes (see Ahdida and Alfonsi 2013, Lemma 27). The operator \mathcal{L}'' is that of

$$Y_t'' = y'' + \sqrt{1 - \sum_{q=1}^n \rho_q^2} \sqrt{x} Z_t,$$

which can be sampled exactly, since it is a Gaussian vector with mean y'' and covariance matrix $(1 - |\rho|^2)tx$. The operator \mathcal{L}_q^c is the infinitesimal generator of

the following SDE:

$$\begin{aligned} Y_t &= y + \rho_q \int_0^t \sqrt{X_s} dW_s g_d^q, \\ X_t &= x + \int_0^t (d-1)\epsilon^2 e_d^q ds + \epsilon \int_0^t \sqrt{X_s} dW_s e_d^q + e_d^q dW_s^T \sqrt{X_s}. \end{aligned} \quad (5.5)$$

Thus, X follows an elementary Wishart process and stays in $\mathcal{S}_d^+(\mathbb{R})$. Using the notation of Ahdida and Alfonsi (2013), X_t follows the law

$$\text{WIS}_d(x, d-1, 0, e_d^q, \epsilon^2 t).$$

Ahdida and Alfonsi (2013, Theorems 9 and 16) give, respectively, an exact and a second (or higher) discretization scheme for this process. We now explain how to calculate Y_t once X_t has been sampled. From (5.5), we have, for $1 \leq i \leq d$,

$$\begin{aligned} d(Y_t)_i &= \rho_q \sum_{j=1}^d (\sqrt{X_t})_{i,j} (dW_t)_{j,q}, \\ d(X_t)_{q,i} &= \epsilon \sum_{j=1}^d (\sqrt{X_t})_{i,j} (dW_t)_{j,q} \\ &\quad + \mathbb{1}_{i=q} \left[(d-1)\epsilon^2 dt + \sum_{j=1}^d (\sqrt{X_t})_{q,j} (dW_t)_{j,q} \right]. \end{aligned}$$

This yields

$$\begin{aligned} (Y_t)_i &= y_i + \frac{\rho_q}{\epsilon} ((X_t)_{q,i} - x_{q,i}) \quad \text{if } i \neq q, \\ (Y_t)_q &= y_i + \frac{\rho_q}{2\epsilon} [(X_t)_{q,q} - x_{q,q} - \epsilon^2(d-1)t]. \end{aligned}$$

Using these formulas together with the exact (respectively, second-order) scheme for X_t , we get an exact (respectively, second-order) scheme for (5.5). By using the composition rules (5.2), we get a second-order scheme for (5.3).

5.2 A faster second-order scheme when $\Omega - \epsilon^2 I_d^n \in \mathcal{S}_d^+(\mathbb{R})$

As explained in Ahdida and Alfonsi (2013), the sampling of each elementary Wishart process in \mathcal{L}_q requires a Cholesky decomposition that has a time complexity of $O(d^3)$. Since the second-order scheme proposed above calls $n \leq d$ times this routine, the whole scheme requires at most $O(d^4)$ operations. However, by adapting an idea that has already been used in Ahdida and Alfonsi (2013) for Wishart processes, it is possible to get a faster scheme if, in addition, we assume that

$$\Omega - \epsilon^2 I_d^n \in \mathcal{S}_d^+(\mathbb{R}).$$

We now present this alternative scheme, which only requires $O(d^3)$ operations.

We consider the splitting $\mathcal{L} = \tilde{\mathcal{L}}' + \tilde{\mathcal{L}}'' + \hat{\mathcal{L}}$ of the operator (5.3), with

$$\begin{aligned}\tilde{\mathcal{L}}' &= \sum_{1 \leq i, j \leq d} (\Omega - \epsilon^2 I_d^n + bx + xb^T)_{i,j} \partial_{x_{i,j}} \\ \hat{\mathcal{L}} &= \sum_{1 \leq i \leq n} d\epsilon^2 \partial_{x_{i,i}} + \frac{1}{2} \sum_{m=1}^d \sum_{1 \leq i, j \leq d} \epsilon [x_{m,i} (I_d^n \rho)_j + x_{m,j} (I_d^n \rho)_i] \partial_{x_{i,j}} \partial_{y_m} \\ &\quad + \frac{\sum_{q=1}^n \rho_q^2}{2} \sum_{m,m'=1}^d x_{m,m'} \partial_{y_m} \partial_{y_{m'}} \\ &\quad + \frac{1}{2} \sum_{1 \leq i, j, k, l \leq d} \epsilon^2 [x_{i,k} (I_d^n)_{j,l} + x_{i,l} (I_d^n)_{j,k} \\ &\quad \quad \quad + x_{j,k} (I_d^n)_{i,l} + x_{j,l} (I_d^n)_{i,k}] \partial_{x_{i,j}} \partial_{x_{k,l}}.\end{aligned}$$

Again, $\tilde{\mathcal{L}}'$ is the operator of the linear ODE

$$x'(t) = \Omega - \epsilon^2 I_d^n + (d-1)\epsilon^2 I_d^n + bx + xb^T;$$

this can be solved exactly and stays in the set of semidefinite positive matrixes by Ahdida and Alfonsi (2013, Lemma 27), since $\Omega - \epsilon^2 I_d^n \in \mathcal{S}_d^+(\mathbb{R})$. We have already seen above that the generator \mathcal{L}'' can be sampled exactly, and we focus now on the sampling of $\hat{\mathcal{L}}$. This relies on the following result.

LEMMA 5.1 *For $x \in \mathcal{S}_d^+(\mathbb{R})$, we consider $c \in \mathcal{M}_d(\mathbb{R})$ such that $c^T c = x$. We define*

$$U_t = c + \epsilon W_t I_d^n, \quad X_t = U_t^T U_t \quad \text{and} \quad Y_t = y + \int_0^t U_s^T dW_s I_d^n \rho.$$

Then, the process (X, Y) has the infinitesimal generator $\hat{\mathcal{L}}$.

PROOF For $1 \leq i, j, m \leq d$, we have

$$d(X_t)_{i,j} = \epsilon \sum_{k=1}^d ((U_t)_{k,i} (dW_t)_{k,j} \mathbb{1}_{j \leq n} + (U_t)_{k,j} (dW_t)_{k,i} \mathbb{1}_{i \leq n}) + \mathbb{1}_{i=j \leq n} d\epsilon^2 dt$$

and

$$d(Y_t)_m = \sum_{k,l=1}^d (U_t)_{k,m} (dW_t)_{k,l} (I_d^n \rho)_l.$$

This leads to

$$\begin{aligned}\langle d(Y_t)_m, d(Y_t)_{m'} \rangle &= \sum_{k,l=1}^d (U_t)_{k,m} (U_t)_{k,m'} (I_d^n \rho)_l^2 dt = \left(\sum_{l=1}^n \rho_l^2 \right) (X_t)_{m,m'} dt, \\ \langle d(Y_t)_m, d(X_t)_{i,j} \rangle &= \epsilon [(I_d^n \rho)_j (X_t)_{m,i} + (I_d^n \rho)_i (X_t)_{m,j}] dt, \\ \langle d(X_t)_{i,j}, d(X_t)_{k,l} \rangle &= \epsilon^2 [(X_t)_{i,k} (I_d^n)_{j,l} + (X_t)_{i,l} (I_d^n)_{j,k} \\ &\quad + (X_t)_{j,k} (I_d^n)_{i,l} + (X_t)_{j,l} (I_d^n)_{i,k}] dt,\end{aligned}$$

which precisely gives the generator $\hat{\mathcal{L}}$. \square

Thanks to Lemma 5.1, it is sufficient to construct a second-order scheme for (U, Y) . Since $\langle d(Y_t)_m, d(U_t)_{i,j} \rangle = \epsilon (U_t)_{i,m} (I_d^n \rho)_j dt$, the infinitesimal generator $\bar{\mathcal{L}}$ of (U, Y) is given by

$$\begin{aligned}\bar{\mathcal{L}} = \frac{\epsilon^2}{2} \sum_{i=1}^d \sum_{j=1}^n \partial_{x_{i,j}}^2 + \frac{\epsilon}{2} \sum_{i,m=1}^d \sum_{j=1}^n \rho_j x_{i,m} \partial_{x_{i,j}} \partial_{y_m} \\ + \frac{\sum_{q=1}^n \rho_q^2}{2} \sum_{m,m'=1}^d (x^T x)_{m,m'} \partial_{y_m} \partial_{y'_{m'}}.\end{aligned}$$

We now use the splitting $\bar{\mathcal{L}} = \sum_{q=1}^n \bar{\mathcal{L}}_q$, with

$$\bar{\mathcal{L}}_q = \frac{\epsilon^2}{2} \sum_{i=1}^d \partial_{x_{i,q}}^2 + \frac{\epsilon}{2} \sum_{i,m=1}^d \rho_q x_{i,m} \partial_{x_{i,q}} \partial_{y_m} + \frac{\rho_q^2}{2} \sum_{m,m'=1}^d (x^T x)_{m,m'} \partial_{y_m} \partial_{y'_{m'}}.$$

By straightforward calculus, we find that $\bar{\mathcal{L}}_q$ is the generator of the following SDE:

$$dY_t = \rho_q U_t^T dW_t g_d^q, \quad dU_t = \epsilon dW_t e_d^q.$$

We note that only the q th row of U is modified. For $1 \leq i \leq d$, we have

$$d(U_t)_{i,q} = \epsilon (dW_t)_{i,q} \quad \text{and} \quad d(Y_t)_m = \rho_q \sum_{j=1}^d (U_t)_{j,m} (dW_t)_{j,q}.$$

This yields

$$\begin{aligned}(Y_t)_m &= (Y_0)_m + \rho_q \sum_{j=1}^d (U_0)_{j,m} (W_t)_{j,q} \quad \text{for } m \neq q, \\ (Y_t)_q &= (Y_0)_q + \rho_q \sum_{j=1}^d (U_0)_{j,q} (W_t)_{j,q} + \frac{\epsilon \rho_q}{2} \sum_{j=1}^d \{(W_t)_{j,q}^2 - t\}.\end{aligned}$$

Using these formulas, we can sample (U_t, Y_t) exactly and then get a second-order scheme for $\hat{\mathcal{L}}$. We note that the simulation cost of $\hat{\mathcal{L}}_q$ requires $O(d)$ operations, while that of $\hat{\mathcal{L}}$ requires $O(d^2)$ operations. Since a matrix multiplication requires $O(d^3)$ operations, this second-order scheme for $\hat{\mathcal{L}}$ and then for \mathcal{L} requires $O(d^3)$ operations, instead of the $O(d^4)$ needed for the scheme described in Section 5.1.

REMARK 5.2 As already mentioned, the dependence between the processes X and Y is the same as that proposed by Da Fonseca *et al* (2008) for a model on asset returns. Therefore, we can use the same splittings as those proposed in Sections 5.1 and 5.2 in order to construct second-order schemes for their model.

5.3 Numerical results

We now turn to the empirical analysis of the convergence of the discretization schemes we have proposed. We will use the following notation.

Scheme 1: the second-order scheme given in Section 5.1, where we use the exact sample of the Wishart part and the exact simulation of the Gaussian variables.

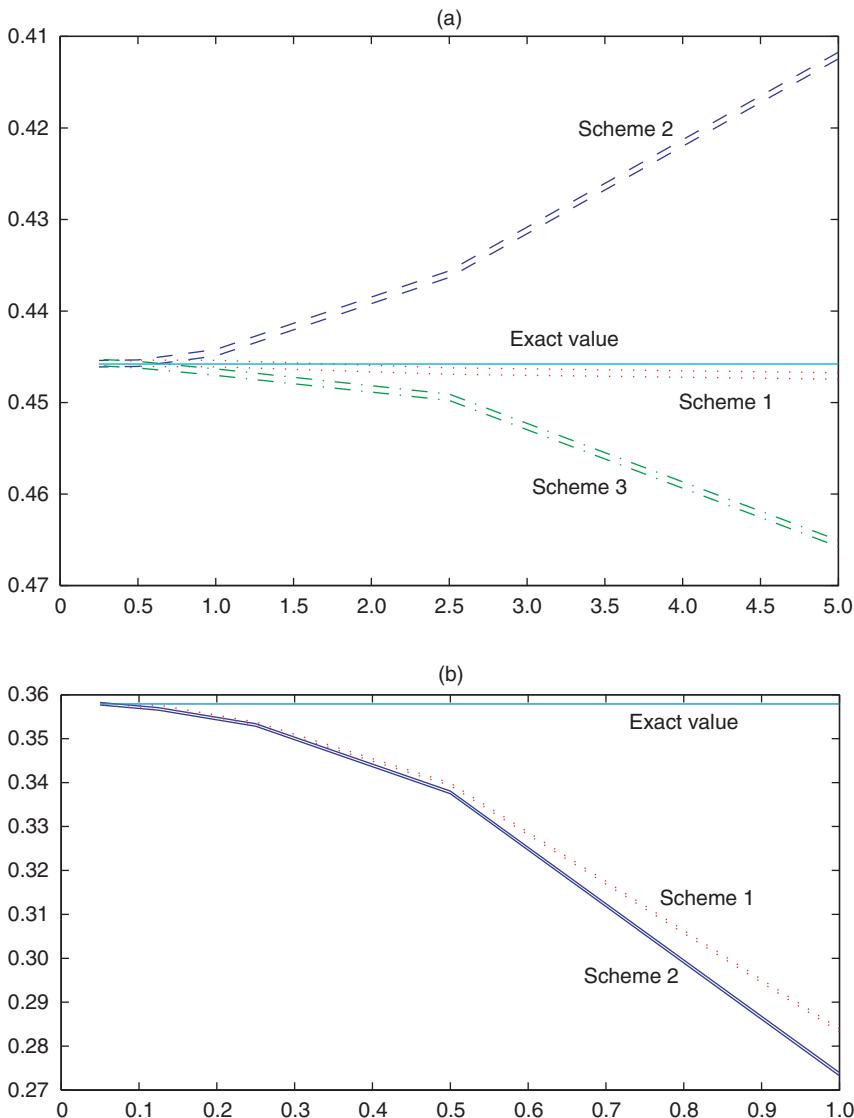
Scheme 2: the second-order scheme given in Section 5.1, where we use the second-order scheme for the Wishart part and replace the simulation of Gaussian variables with random variables that match the five first moments (see Ahdida and Alfonsi 2013, Theorem 16; Equation (36)).

Scheme 3: the second-order scheme given in Section 5.2.

In order to assess whether the potential second-order schemes we have proposed for \mathcal{L} do indeed give a weak error of order 2, we start by analyzing the weak error for quantities that we can compute analytically. Namely, we consider $\mathbb{E}[\exp(-i(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T))]$, which can be calculated by solving a system of differential equations similar to (3.10). We then compare the values obtained by Monte Carlo simulation and the value obtained by solving the system of differential equations. As shown in parts (a) and (b) of Figure 5, we observe a weak error that is compatible with the rate of $O(1/N^2)$. When it is well defined, Scheme 3 is preferred, since it is much faster than the others.

6 COMPARISON OF THE DIFFERENT NUMERICAL METHODS

The goal of this section is to compare the computational time needed to price vanilla instruments in the model using the different numerical methods. We consider the case of a 6M × 1Y caplet with strike 1%, which means $T = 1$ and $\delta = 1/2$. Its price is

FIGURE 5 Weak error convergence.

Parameters: $p = d = 3$, 10^7 Monte Carlo samples, $T = 5$. The real value of $\mathbb{E}[\exp(-i(\text{Tr}(\Gamma X_T) + \Lambda^T Y_T))]$ as a function of the time step T/N . (a) $\Gamma = 0.05I_d$, $\Lambda = 0.021_d$ and the diffusion parameters $x = 0.4I_d$, $y = 0.21_d$, $\Omega = 2.5I_d$, $n = d$, $\rho = 0$, $b = 0$, $\kappa = 0$, $c = I_d$. The value obtained by solving the ODE is -0.445787 . (b) $\Gamma = 0.2I_d + 0.04q$, $\Lambda = 0.21_d$ and the diffusion parameters $x = 0.4I_d + 0.2q$, $y = 0.21_d$, $\Omega = 0.5I_d$, $n = d$, $\rho = -0.31_d$, $b = -0.5I_d$, $\kappa = 0.1I_d$, $c = I_d$, where $q_{i,j} = \mathbf{1}_{i \neq j}$. The value obtained by solving the ODE is 0.357901 . For each scheme, the two curves represent the upper and lower bound of the 95% confidence interval.

given by

$$\begin{aligned} \frac{1}{\delta} \mathbb{E} \left[\exp \left(- \int_0^T r_s \, ds \right) (1 - (1 + K\delta) P_{T,T+\delta})^+ \right] \\ = \frac{P_{0,T}}{\delta} \mathbb{E}^T [(1 - (1 + K\delta) P_{T,T+\delta})^+]. \end{aligned}$$

We will compare the expansion and the Monte Carlo method with respect to the Fourier inversion method presented in Carr and Madan (1999) and Lee (2004). Their approach can be directly applied to caplets by working with the forward caplet price. Let us note that this method can be adapted for swaptions by making the same approximation that we used for the expansion (see Schrager and Pelsser 2006; Singleton and Umantsev 2002). Here, we consider four numerical methods.

- The Monte Carlo method. This consists of using the second-order scheme for (X, Y) with a time step of $1/8$ and 10 000 paths in order to approximate $(1/\delta) \mathbb{E}[\exp(-\int_0^T r_s \, ds)(1 - (1 + K\delta) P_{T,T+\delta})^+]$.
- The expansion up to order 2. The integrals that define the coefficients c_i , d_i and e_i are approximated using a trapezoidal rule and a time step of $1/20$.
- The Fourier transform under \mathbb{P}^T . Starting from the expectation under the T -forward measure, we use the construction of Carr and Madan (1999). In Carr and Madan (1999, Equation (5)), we use $\alpha = 1.25$, truncate the integral at 375 and use a Simpson's rule with a discretization step of $1/8$. Since we calculate only one price here, we do not use the fast Fourier transform, which would have generated further constraints between the discretization and strike grids.
- The Fourier transform under $\mathbb{P}^{T+\delta}$. This is the same method as the previous one, but we use the pricing formula (4.1) under the $T + \delta$ forward measure. We use the same parameters to approximate the integral, and again $\alpha = 1.25$.

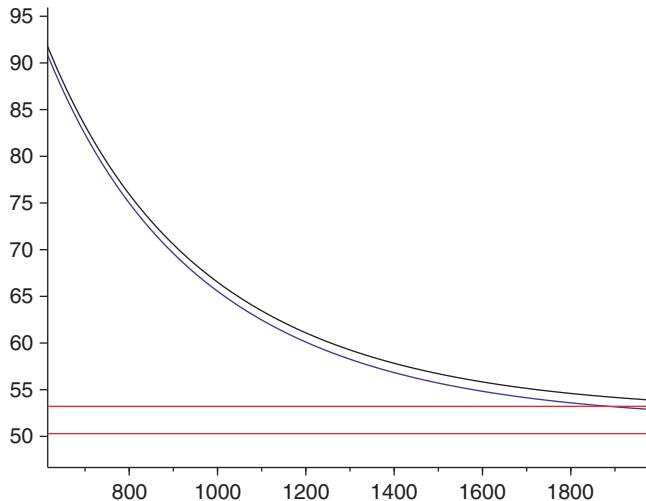
For each method, we have indicated in Table 1 the accuracy and the computation time. A striking fact is that the method based on the Fourier transform is not so efficient in this context, even though the Fourier inversion is in dimension one. The reason for this is that the evaluation of the Fourier transform requires us to numerically solve matrix Riccati differential equations, for which we take a time step of $1/8$. Figure 6 indicates that, in our case, a minimum of 2000 evaluations is necessary to get precision similar to that of the Monte Carlo method. Thus, a basic application of the method of Carr and Madan is not very efficient: the bottleneck is finding a smarter way to calculate the characteristic function. In comparison, the Monte Carlo method is not much more time consuming and allows us to calculate the price for all strikes and maturities at the same time. Lastly, we observe that the expansion method is much

TABLE 1 Price of the $6M \times 1Y$ caplet with strike 1% using different methods, with parameter set (4.3) and $\rho = (-0.4, -0.2)$.

Pricing method	Price (bp)	CPU time (s)
Monte Carlo price	51.75 ± 1.46 (95% CI)	43.3
Expansion	52.33	0.686
Fourier under P^T	53.84	31.6
Fourier under $P^{T+\delta}$	52.87	33.6

Computations are made on a personal laptop with 4GB RAM and a 2.13GHz CPU.

FIGURE 6 Numerical analysis of the truncation error in the pricing by Fourier transform.



Convergence of the Fourier transform price of the $6M \times 1Y$ caplet, with strike 1% and a fixed time step of $1/8$, in function of the number of discretization steps n_{st} . The integration (see Carr and Madan 1999, Equation (5)) is thus made on $[0, n_{st}/8]$. The parallel lines indicate the 95% confidence interval obtained by Monte Carlo simulation.

faster than the others, but it is limited to short maturities, as indicated in Section 4.3. It can therefore be used as a tool for quickly calibrating the model to some key features, such as the at-the-money price and skew.

7 CONCLUSION

The contribution of our paper is twofold. First, we aimed to define a Wishart-driven ATSM for interest rates, in which the parameters and state variables of the model allowed a clear interpretation in terms of the yield curve dynamics, and to provide an

efficient numerical framework for implementing the model. Other ATSMs involving Wishart processes have been proposed, for example, by Bensussan (2010) and Gnoatto (2012). A pitfall of general ATSMs is offering an abundant parameterization with few intuitions for the practitioner. Here, we believe that presenting the model as a perturbation of the standard LGM is a good way of getting to grips with it, of having a better understanding of the parameters and of having a starting point for the calibration procedure. Let us mention here that getting a reliable and stable calibration procedure of the model is beyond the scope of this paper. In particular, the choice of the dimensions p and d should be discussed with regard to real data. Also, we have made the choice in this paper to present the model with constant (as opposed to time-dependent) parameters: only the factors are meant to describe the state of the interest rate market. Thus, this version of the model has a priori a limited flexibility to calibrate to the swaption volatility cube compared with a fully non-homogeneous term structure model with time-dependent parameters, such as the stochastic volatility forward Libor model of Piterbarg (2003) and the stochastic volatility Cheyette model considered by Andreasen (2005). A full discussion on the calibration of our model as well as its comparison to other models is left for future research.

The other contribution of this paper is to investigate different numerical methods for the model. We know that having efficient numerical methods is a prerequisite of using a model. In addition, our results could be interesting for other models based on Wishart dynamics. As the state variable dynamics are affine, their Fourier and Laplace transforms are tractable and can be obtained by solving ODEs. Therefore, Fourier transform pricing methods can be applied to price vanilla interest rate options in the model. However, the results of our numerical investigation suggest that standard Fourier-based pricing methods suffer from numerical inefficiency. This is due to a rather lengthy evaluation of the characteristic function together with a slow convergence rate of the Fourier transform discretization. A smarter way to evaluate the characteristic function and solve the corresponding differential equation has to be investigated in order to make this method more attractive. As an alternative, we have developed a pricing method for vanilla interest rate options based on a perturbation of the infinitesimal generator of the state variables. This method provides a fast pricing tool for the products that would typically be used for model calibration. The method is particularly efficient for short expiries, but it proves to have limitations for long-dated options. Also, the expansion provides analytical expressions for the implied volatility of caplets and swaptions. This is important in confirming our intuition with regard to the role of the parameters, and it can be used to initialize the calibration routine. Lastly, we propose a second-order discretization scheme for the model, which is useful in running the Monte Carlo method. This scheme is easy to implement and very efficient in practice. In addition, it can be adapted easily to a wider range of financial models that use the same dependence structure between the vector Y and its instantaneous

Wishart covariance matrix X , such as the Wishart affine stochastic correlation model developed by Da Fonseca *et al* (2008, 2014). Moreover, it is, as far as we know, the first second-order discretization scheme that can handle this instantaneous covariance structure.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the writing of this work.

ACKNOWLEDGEMENTS

Most of this work was done when Ernesto Palidda was working for the Groupe de Recherche Opérationnelle of Crédit Agricole. We thank Christophe Michel, Victor Reutenauer, Anas Benabid and Nicole El Karoui for interesting and fruitful discussions about this paper. This research also benefited from the support of the “Chaire Risques Financiers”, Fondation du Risque.

REFERENCES

- Ahdida, A., and Alfonsi, A. (2013). Exact and high-order discretization schemes for Wishart processes and their affine extensions. *Annals of Applied Probability* **23**(3), 1025–1073 (<http://doi.org/btvj>).
- Alfonsi, A. (2010). High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Mathematics of Computation* **79**(269), 209–237 (<http://doi.org/b5q5f2>).
- Andersen, L., and Piterbarg, V. (2010). *Interest Rate Modeling, Volume 2*. Atlantic Financial Press.
- Andreasen, J. (2005). Back to the future. *Risk* **18**(9), 104–109.
- Benabid, A., Bensussan, H., and El Karoui, N. (2008). Wishart stochastic volatility: asymptotic smile and numerical framework. Working Paper, Centre de Mathématiques Appliquées, Ecole Polytechnique. URL: <http://bit.ly/2fvrFb6>.
- Bensussan, H. (2010). Interest rate and longevity risks: dynamic modelling and applications to derivative products and life insurance. PhD Thesis, Centre de Mathématiques Appliquées, Ecole Polytechnique. URL: <http://bit.ly/2gECCJf>.
- Bergomi, L., and Guyon, J. (2012). Stochastic volatilities orderly smiles. *Risk* **25**(5), 60–66.
- Brigo, D., and Mercurio, F. (2006). *Interest Rate Models: Theory and Practice: With Smile, Inflation and Credit*, 2nd edn. Springer Finance.
- Carr, P., and Madan, D. (1999). Option valuation using fast Fourier transform. *The Journal of Computational Finance* **2**(4), 61–73.
- Collin-Dufresne, P., and Goldstein, R. (2002). Pricing swaptions within an affine framework. *Journal of Derivatives* **10**(1), 9–26.
- Cox, J. C., Ingersoll, J. E., Jr., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**(2), 385–407 (<http://doi.org/cbb2pm>).

- Cuchiero, C., Filipović, D., Mayerhofer, E., and Teichmann, J. (2011). Affine processes on positive semidefinite matrices. *Annals of Applied Probability* **21**(2), 397–463 (<http://doi.org/d5384c>).
- Da Fonseca, J., Grasselli, M., and Tebaldi, C. (2008). Option pricing when correlations are stochastic: an analytical framework. *Review of Derivatives Research* **10**(2), 151–180.
- Da Fonseca, J., Grasselli, M., and Ielpo, F. (2014). Estimating the Wishart affine stochastic correlation model using the empirical characteristic function. *Studies in Nonlinear Dynamics and Econometrics Journal* **18**(3), 253–289.
- Dai, Q., and Singleton, K. J. (2000). Specification analysis of affine term structure models. *Journal of Finance* **55**(5), 1943–1978 (<http://doi.org/d2hxrx>).
- D'Aspremont, A. (2003). Interest rate model calibration using semidefinite programming. *Applied Mathematical Finance* **10**(3), 183–213 (<http://doi.org/fn3hhk>).
- Dieci, L., and Eirola, T. (1994). Positive definiteness in the numerical solution of Riccati differential equations. *Numerische Mathematik* **67**(3), 303–313 (<http://doi.org/ftwm57>).
- Duffie, D., and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance* **6**, 379–406.
- Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability* **13**(3), 984–1053 (<http://doi.org/bkw26k>).
- El Karoui, N., and Lacoste, V. (1992). Multifactor models of the term structure of interest rates. Preprint.
- El Karoui, N., Lepage, C., Myneni, R., Roseau, N., and Viswanathan, R. (1991). The valuation and hedging of contingent claims with Markovian interest rates. Preprint.
- Filipović, D. (2009). *Term-Structure Models: A Graduate Course*. Springer Finance (<http://doi.org/d4w22k>).
- Fouque, J.-P., Papanicolaou, G., and Sircar, K. R. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, Cambridge.
- Geman, H., El Karoui, N., and Rochet, J.-C. (1995). Changes of numéraire, changes of probability measure and option pricing. *Journal of Applied Probability* **32**(2), 443–458.
- Gnoatto, A. (2012). The Wishart short rate model. *International Journal of Theoretical and Applied Finance* **15**(8), 1250056 (<http://doi.org/btvk>).
- Lee, R. W. (2004). Option pricing by transform methods: extensions, unification, and error control. *The Journal of Computational Finance* **7**(3), 51–86.
- Palidda, E. (2015). Managing interest rates derivatives with stochastic variance–covariance. PhD Thesis, École des Ponts ParisTech. URL: <http://bit.ly/2gkq7hz>.
- Piterbarg, V. (2003). A Stochastic volatility forward Libor model with a term structure of volatility smiles. Working Paper, Social Science Research Network. URL: <http://ssrn.com/abstract=472061>.
- Piterbarg, V. (2009). Rates squared. *Risk* **22**(1), 100–105.
- Rydberg, T. H. (1997). A note on the existence of unique equivalent martingale measures in a Markovian setting. *Finance and Stochastics* **1**(3), 251–257 (<http://doi.org/cwjdz7>).
- Schrager, D. F., and Pelsser, A. A. J. (2006). Pricing swaptions and coupon bond options in affine term structure models. *Mathematical Finance* **16**(4), 673–694 (<http://doi.org/bdrqkt>).
- Singleton, K. J., and Umantsev, L. (2002). Pricing coupon-bond options and swaptions in affine term structure models. *Mathematical Finance* **12**(4), 427–446 (<http://doi.org/dw4r75>).

- Tanaka, K., Yamada, T., and Watanabe, T. (2010). Applications of Gram–Charlier expansion and bond moments for pricing of interest rates and credit risk. *Quantitative Finance* **10**(6), 645–662 (<http://doi.org/fhncd4>).
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**(2), 177–188 (<http://doi.org/fggzb2>).

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net