Efficient Pricing and Estimation Methods in Finance

Lorenzo Pitotti

Imperial College Business School

Imperial College London

Submitted to Imperial College London

In partial fulfilment of the requirement for the degree of

Doctor of Philosophy

February 2013

To my parents Maria Teresa Scherillo and Giuseppe Pitotti.

Acknowledgements

I am extremely grateful to my supervisor Professor Enrico Biffis for his continuous support throughout many years and precious guidance. I would also like to thank Serguei Vassilovski, Raymond Lee and Richard Black of Algorithmics, an IBM Company, for the very helpful feedback and discussions which have been critical to achieving the results of Part II of this thesis, as well as Mourad Benouarets, Martin Botha and John Kougioumtzis for comments and suggestions. I am also thankful for the several useful discussions with professors and colleagues at Imperial College London, in particular Paolo Zaffaroni, Filip Zikes, Karim Abadir, Aleksandar Mijatović and Martijn Pistorius and also Roberto Renò of the university of Siena. Last but not least I am grateful to Wenting Fan for her encouragement and patience throughout this journey.

Contents

List of Figures

Part I

Predictable Representation Property - Applications to Finance

In Chapter 1 we introduce a new methodology for estimating upper bounds for the value of Bermudan products that does not require resorting to nested simulation and is applicable when the underlying variables are driven by Lévy processes. The suggested approach generalizes Belomestny et al. [2009], which is limited to the Brownian setting, by estimating the optimal Lagrangian martingale in the dual representation of the price of a Bermudan product taking advantage of the predictable representation property for Lévy processes introduced by Nualart and Schoutens [2000]. In addition, we present an easily implementable and computationally efficient variant of the methodology that applies when the value of the underlying financial variables can be expressed using non path-dependent functions of Lévy processes (as in the most common linear and exponential Lévy models) and requires the simulation of variables only at the exercise dates of the Bermudan product.

Chapter 1

Upper Bounds in a Lévy Setting without Nested Simulation

1.1 Introduction

The past decade has seen the introduction and widespread adoption of new methods for the pricing of financial products with American and Bermudan exercise using Monte Carlo methods. Until then, Monte Carlo methods were widely seen as unsuitable to be applied in this context because of the so called Monte Carlo on Monte Carlo problem: the fact that a straight-forward application of Monte Carlo to "optimal stopping" problems requires the recursive nesting of simulated paths at each exercise time for estimating the continuation value of the option, resulting in an infeasible number of simulations. For this reason, the preferred approaches for the pricing of American and Bermudan contracts had traditionally been the use of binomial (Cox et al. [1979]) and trinomial (Boyle [1986]) trees and finite difference methods (e.g. Brennan and Schwartz [1977]), whose lattice structure and backward calculations make them natural candidates for dealing with optimal stopping problems. However, these methods have the drawback of being computationally very demanding when the dimensionality of the problem increases (particularly above three dimensions). The first Monte Carlo method for pricing American and Bermudan options that has seen wide adoption is the Least-Squares Monte Carlo (LSM), popularized by Longstaff and Schwartz [2001]

whose main ideas had been previously introduced, in different forms, by Carriere [1996] and in Tsitsiklis and Van Roy [1999] and Tsitsiklis and Van Roy [2001]. LSM has been proven (see Clément et al. [2002] and Stentoft [2004b]) to converge to the true option price, however its estimates are, by construction, biased from below (i.e. they represent a "lower bound" to the option price). In the absence of an unbiased method, this left open the need for calculating a corresponding "upper bound" allowing to bracket the true value within a confidence interval. An answer to this problem was given by the duality methods of Rogers [2002] and Haugh and Kogan [2004] which used ideas already in Davis and Karatzas [1994]. At present, the most popular primal-dual Monte Carlo approach for the pricing of Bermudan products is the one proposed by Andersen and Broadie [2004], which merges LSM and duality approach in an integrated method providing both lower and upper bounds. Their method can achieve tight bounds but it requires considerably more intensive computations for the calculation of the upper bounds than it needs for the lower bound. The reason for this is that the algorithm uses nested simulations for estimating the optimal Lagrangian martingale in the dual representation of the price. In an attempt to solve this problem, Belomestny et al. [2009] and Schoenmakers et al. [2012] (among others) have recently proposed methods for the calculation of upper bounds that do not require nested simulation. In this chapter we extend and generalize the results of Belomestny et al. [2009], which are limited to the Brownian setting, to Lévy processes by taking advantage of the Predictable Representation Property for Lévy processes introduced by Nualart and Schoutens [2000]. In addition, we present an easily implementable and computationally efficient variant of the methodology that applies when the value of the underlying financial variables can be expressed using non path-dependent functions of Lévy processes (as in the most common linear and exponential Lévy models) and requires the simulation of variables only at the exercise dates of the Bermudan product. In sections 1.2 and 1.3 we discuss the problem of pricing instruments with American and Bermudan exercise and review the approaches available in the literature for pricing these types of products with Monte Carlo methods while in Section 1.4 we illustrate the main results of Nualart and Schoutens [2000] on the predictable representation of Lévy processes; in the following two sections we present the main results of this chapter, first in

a general setting then in the context of non path-dependent models. Finally, in Section 1.7 we present the results of numerical examples which show the proposed non-nested estimator achieving tight upper bounds in a computationally efficient way. Section 1.8 summarizes our conclusions.

1.2 American Options

What distinguishes American from European options is the specific way in which the holder can exercise the right embedded in the contract. More specifically, let X_t be the price at time t of the underlying asset of the option, T be the final exercise date of the contract and Φ, $\Phi(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$, its discounted payoff function (we assume that all payments are discounted to time 0 using $\mathcal{N}_t = \exp\left(\int_0^t r_u du\right)$ as *numéraire* (see Shreve [2004]) where r is the instantaneous risk-free interest rate). In the case of an European option the amount $\Phi(T, X_T)\mathcal{N}_T$ is paid to the holder at time T. On the other hand, the holder of an American option has the right to exercise the contract and obtain the amount $\Phi(t, X_t)\mathcal{N}_t$ at any time t prior to T. The situation is complicated by the fact that the exercise time t does not have to be chosen *a priori* and may be chosen by the holder on the basis of the information available at time t. The time of exercise is therefore a random variable itself and its mathematical formulation is given in terms of *stopping times* (see Shreve [2004]).

1.2.1 An Optimal Stopping Problem

The fundamental theorems of asset pricing (Shreve [2004]) tell us that arbitrage-free prices of contingent claims can essentially be obtained from conditional expectations of their discounted future payoffs under a risk-neutral probability measure. More formally, let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P\right)$ be a filtered probability space and Q a (risk-neutral) equivalent martingale measure associated with numéraire $\mathcal{N}$. Then, making the additional assumption of market completeness (i.e. the risk-neutral probability measure is unique), we can express the time t value of any contingent future (discounted) payoff Ψ_T, paid at time T, as $V_t = \mathbb{E}_t^Q[\Psi_T]$, where $\mathbb{E}_t^Q[\cdot]$ denotes the conditional expectation, under measure Q, with respect to $\mathcal{F}_t$ (i.e.

the information available at time t). In the case of American options, a no-arbitrage argument (Shreve [2004]) implies that finding the value of the contract is equivalent to solving an *optimal stopping problem*, whose solution is given by the exercise strategy (the optimal stopping time) that maximizes the expected future payoff to the holder of the option. In mathematical terms:

$$V_0 = \sup_{\tau} \left[\mathbb{E}_0^Q \left[\Phi(\tau, X_\tau) \right] \right] \ , \tag{1.2.1}$$

where X is a Markov process adapted to $(\mathcal{F}_t)_{t\geq 0}$ and τ is allowed to vary over the class of stopping times, i.e. τ is a random variable with values in $\mathbb{R}_0^+$ such that $\{\tau \leq t\} \in \mathcal{F}_t$. The optimization problem on the right hand side of (1.2.1) is solved by the stopping time

$$\tau_0^* = \inf \{ t \geq 0 : X_t \in B \} \ , \tag{1.2.2}$$

for some optimal exercise region B. The stopping time τ_0^* can be referred to as a *rational exercise policy*, in the sense that it maximizes the initial arbitrage-free value of the resulting claim. Problems of this kind are quite hard to solve, and analytically they lead to so called "free boundary value problems" instead of the corresponding parabolic PDEs for the European counterparts. Practically, for American contracts, very few analytical formulas are at hand[1]. The value process $\widetilde{V} = \left(\widetilde{V}_t\right)_{t\geq 0}$ of the American option, where $\widetilde{V}_0 = V_0$, is referred to as the *Snell envelope* of Φ. In the discrete time case (for the continuous time case see Karatzas and Shreve [1998]) the Snell envelope $\widetilde{V}$ of the process $\Phi = (\Phi(T_n, X_n))_{n=0,\ldots,N}$ is defined as follows:

$$\begin{cases} \widetilde{V}_N = \Phi(T_N, X_N) \\ \widetilde{V}_n = \max\left(\Phi(T_n, X_n), \mathbb{E}_{T_n}\left[\widetilde{V}_{n+1} \right] \right) \end{cases} . \tag{1.2.3}$$

[1] One special case where the price of an American contract is very easy to calculate is when we consider an American call option on a non dividend paying underlying stock. In this case, as it is commonly known, simple arbitrage reasoning demonstrates that the option price will be equivalent to the price of an European call option having the same maturity date and strike price. For American call options with discrete dividends, the argument above can be extended to show that it can only be optimal to exercise the option either at the final time T or at one of the dividend times.

This formulation underlines the *backwards* nature of the valuation problem and shows that the problem of finding the optimal decision rule is equivalent to that of finding the best estimate of the conditional expectations, under the pricing measure, of the discounted payoffs from continuation (from not exercising the option at the present time). It is, in fact, the knowledge of such conditional expectation that may allow the owner of the contract to make an informed decision on whether to exercise or keep holding the option.

1.3 Pricing by Simulation

1.3.1 Lower Bounds

A straightforward application of Monte Carlo methods to (1.2.1) is impractical because one cannot run a Monte Carlo simulation for every possible exercise strategy. However, once an exercise strategy is chosen, valuation becomes simple: at each time step along the simulated paths, we can check the exercise strategy and, if it says "exercise" we return the exercised value at that time step suitably discounted back to time 0, otherwise we proceed to the next step. The main issue is to determine which one is the best exercise strategy, noting that as long as the chosen strategy does not coincide with the optimal τ_0^*, the Monte Carlo estimate will only represent a 'lower bound' to the option price. The central contribution of the LSM approach (described below) has been to provide a convenient way to estimate the optimal exercise strategy τ_0^* using a single set of Monte Carlo paths.

The Least-squares Monte Carlo approach

The Least-squares Monte Carlo approach (LSM) is a simple yet powerful method for approximating the value of American options that was introduced by Longstaff and Schwartz [2001]. Their work developed and simplified an idea previously analysed by Carriere [1996] and by Tsitsiklis and Van Roy [1999] and Tsitsiklis and Van Roy [2001]: estimating the conditional expectation of the payoff from keeping the option alive at each possible exercise point from a simple least square regression using the cross-sectional information provided by Monte Carlo simulation. In the regression process they use a set of *basis functions* in the

underlying asset prices. The fitted values are then taken as the expected continuation values. Comparing these estimations with the immediate exercise values, they identify the optimal exercise decision. The procedure is repeated recursively going back in time. By estimating the conditional expectation function for each exercise date, they obtain a complete specification of the optimal exercise strategy along each path. The advantage of this approach is that it is readily applicable in path-dependent and multifactor situations where traditional finite difference techniques cannot be used; furthermore, as it was showed in Stentoft [2004a], the LSM method, when compared to the traditional binomial tree approach, displays a much better trade-off between computational time and precision as the number of stochastic factors is increased.

More formally, let's assume a finite time horizon, $[0, T]$, in which we define, as before, a filtered probability space, $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P\right)$, and an equivalent martingale measure Q. Let $\Phi(\tau^*_{t^+}, X_{\tau^*_{t^+}})$ denote the discounted option cash-flow, conditional on the option being exercised after t and the option holder following the optimal stopping strategy at every time after t, i.e. $\tau^*_{t^+} = \inf\{u > t : X_u \in B\}$. The American option is approximated by its Bermuda counterpart, assuming a finite number of exercise dates $\mathcal{E} = \{T_1, ..., T_{\mathcal{I}}\}$, where $0 < T_1 < T_2 < ... < T_{\mathcal{I}} = T$. For every $T_i \in \mathcal{E}$, the time-T_i continuation value F_{T_i}, $F_{T_i} : \Omega \to \mathbb{R}$, is, under no-arbitrage conditions, the risk-neutral expectation of the future discounted cash-flows $\Phi(\tau^*_{T_i^+}, X_{\tau^*_{T_i^+}})$ conditional upon the information available at time T_i:

$$F_{T_i} = \mathbb{E}^Q_{T_i}\left[\Phi\left(\tau^*_{T_i^+}, X_{\tau^*_{T_i^+}}\right)\right] \ . \tag{1.3.1}$$

Longstaff and Schwartz assume that the unknown functional F_{T_i} can be approximated as a linear combination of a finite set of basis functions in the value X of the asset underlying the option. We can write this as:

$$F_{T_i} \approx F^M_{T_i} = \sum_{m=0}^{M-1} \phi_m\left(X_{T_i}\right) a_m\left(T_i\right) \ , \tag{1.3.2}$$

where $\{\phi_m(.)\}_{m=0}^{M-1}$ is a finite set of functions from a suitable basis. The regression based algorithm on the simulated paths estimates the value of the coefficients

$\{a_m(.)\}_{m=0}^{M-1}$ and provides in turn a very convenient and powerful way for estimating F_{T_i}.

The convergence properties of LSM have been studied by Clément et al. [2002], who proved that, as the number of basis functions goes to infinity, the estimated conditional expectation ($F_{T_i}^M$) tends to the true conditional expectation (F_{T_i}) and showed that the normalized estimation error is asymptotically Gaussian. Stentoft [2004b] has further analysed the asymptotic properties of LSM and proved the mean squared convergence of $F_{T_i}^M$ to F_{T_i} (in the two-period case), providing the rate of convergence.

1.3.2 Upper Bounds - The Duality Approach

As we mentioned above, the sub-optimality of the approximations to τ_0^* obtained using methods such as LSM, result in estimates of the option price that are biased from below. In other words, the Monte Carlo estimator will converge to a lower bound to the option price rather than the option price itself. In the absence of fully unbiased estimators, this situation leaves open the need to determine corresponding 'upper bounds' to the option price that may allow bracketing the unknown true value within an interval. Rogers [2002] and Haugh and Kogan [2004], using ideas that appeared previously in Davis and Karatzas [1994], introduced a dual representation of the price of American options which, rather than expressing its value as a maximization problem (as in (1.2.1)), obtains it in the form of a minimization, therefore providing a way to estimate upper bounds to the value of the contract. In particular, the option value is given by:

$$V_0 = \inf_{M \in \Pi} \mathbb{E}_0^Q \left[\max_t \left(\Phi(t, X_t) - M_t \right) \right] \tag{1.3.3}$$

where Π is the space of adapted martingales for which $\sup_{0 \le t \le T} |M_t| \in L^1$. The value of the American option is achieved at M^* which can be interpreted as the martingale part of the Doob decomposition of the Snell envelope V_t (see Williams [1991]). The proof of the equivalence of the solutions of the primal (1.2.1) and dual (1.3.3) problems can be found in Haugh and Kogan [2004].

Nested Methods

Andersen and Broadie [2004] have provided a very popular methodology for obtaining both lower and upper bounds to the Bermudan option price by integrating the algorithm of Longstaff and Schwartz [2001] with a 'nested' Monte Carlo method for estimating the upper bound. Their approach has the benefit of being fairly easy to implement, however it has the disadvantage that the nested Monte Carlo simulations involved in the estimation of the upper bound require much longer computation time than it is required for estimating the lower bound (up to 15 times more in their tests), particularly as the number of exercise opportunities increases. When applying their pricing algorithm, one has to first determine an estimation of the optimal exercise strategy using the LSM method; then a routine is called repeatedly to generate an upper bound, that complements the lower bound, consistent with the proposed exercise strategy. One significant difference to the algorithm of Haugh and Kogan [2004] is that they do not build or require an approximation to the option price process throughout the state space. Instead, the algorithm uses only the information from the approximation to the optimal exercise strategy, which has benefits in terms of computation time and approximation error. On the other hand, as we mentioned above, their procedure remains computationally intensive because of the path-wise nested Monte Carlo simulation required to estimate the evolution of the optimal martingale M^* in equation (1.3.3) (see Figure 1.1 for a visual illustration of the impact of nesting).

Upper bounds without nested Monte Carlo

Belomestny et al. [2009] introduced a new method, limited to the Brownian Motion setting, which allows the estimation of the optimal martingale M^* in (1.3.3) through a single, non-nested, Monte Carlo experiment. This approach uses the predictable representation property of Brownian Motion to derive a regression estimator for M^* which can also be used as a control variate for the nested primal-dual estimator of Andersen and Broadie. In a recent contribution, Schoenmakers et al. [2012] propose another non-nested approach for the estimation of upper bounds to the price of Bermundan products which has the peculiarity that no

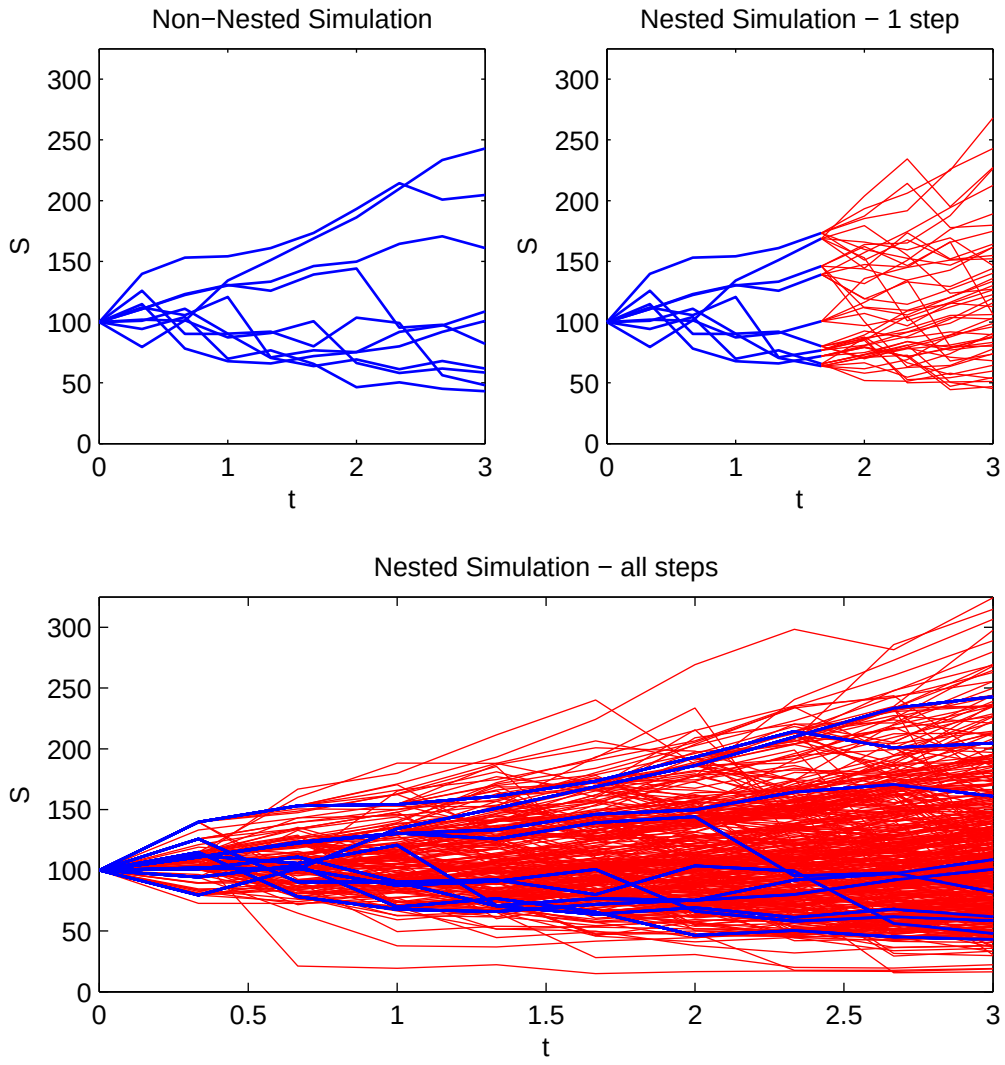


Figure 1.1: Example of nested simulation (10 paths, 5 nested paths)

approximation to the Snell envelope is involved in the calculations and it can therefore be considered a 'pure dual' method. One aspect that Belomestny et al. [2009] and Schoenmakers et al. [2012] have in common is that their implementation algorithms involve knowledge of the analytical formula for the discounted price of the corresponding European options and the application of Malliavin calculus techniques to obtain *ad-hoc* sets of basis functions, which makes implementation more complex and represents a limit to the general applicability of their approaches.

1.4 Predictable Representation and Orthogonal Polynomials

The Predictable (or Martingale) Representation property has been known to hold only for Brownian motion and the Poisson process. However, following the seminal work of Nualart and Schoutens [2000], new literature has recently appeared providing similar results for more general classes of stochastic processes. In particular, Nualart and Schoutens [2000] (see also Jamshidian [2005]) establish a predictable representation property that applies to the class of Lévy processes. Their results are achieved by introducing the compensated power jump processes, also known as *Teugels* martingales, that, as we will explain in greater detail in paragraph 1.4.2, is intrinsically related to the theory of orthogonal polynomials.

1.4.1 The Predictable Representation Property

As we have already anticipated, Nualart and Schoutens [2000] were able to establish a martingale representation property also for the more general case of a filtration generated by a Lévy process L satisfying an exponentially decaying condition on the Lévy measure. More specifically, they showed that every square integrable martingale G with respect to the filtration generated by L, admits a representation as an infinite sum of the form

$$G_t = G_0 + \sum_{i=1}^{\infty} \int_0^t \phi_s^{(i)} dH_s^{(i)} \tag{1.4.1}$$

where $\phi^{(i)}$, $i = 1, 2, \ldots$, are predictable processes , and $\left\{H^{(i)}, i = 1, 2, \ldots\right\}$ is a set of pairwise strongly orthogonal martingales obtained as linear combinations of Teugels martingales (as defined in section 1.4.2 below).
In a more recent contribution, Jamshidian [2005], applying the concepts and techniques introduced by Nualart and Schoutens, together with a general strong orthogonalization procedure presented by Davis and Karatzas [1994], has proposed a generalization of the predictable representation property to processes where the (generalized) Lévy measure is allowed to be a stochastic process adapted to a Brownian filtration. One example of processes included in this class are Cox processes, also referred to as Poisson processes with stochastic intensity.

1.4.2 Teugels Martingales

Consider a Lévy process L with Lévy triplet given by (a, σ, ν), the *Teugels* martingale of order i of L is defined as :

$$Y_t^{(i)} := L_t^{(i)} - \mathbb{E}\left[L_t^{(i)}\right] \tag{1.4.2}$$

where

$$L_t^{(i)} = \sum_{0<s\leq t} \left(\Delta L_s\right)^i \tag{1.4.3}$$

and

$$\mathbb{E}\left[L_t^{(i)}\right] = \mathbb{E}\left[\sum_{0<s\leq t} \left(\Delta L_s\right)^i\right] = t\int_{-\infty}^{\infty} x^i \nu\left(dx\right) \ . \tag{1.4.4}$$

Let S_1 denote the space of all real polynomials defined on the positive real line with scalar product $\langle ., . \rangle_1$ defined by

$$\begin{aligned} \left\langle p_n\left(x;\delta_1\right), p_m\left(x;\delta_2\right)\right\rangle_1 &= \int_{-\infty}^{\infty} p_n\left(x;\delta_1\right) p_m\left(x;\delta_2\right) x^2 \nu\left(dx\right) &(1.4.5)\\ &\quad + \sigma^2 p_n\left(0;\delta_1\right) p_m\left(0;\delta_2\right) &(1.4.6)\end{aligned}$$

and let S_2 denote the space of all linear transformations of the Teugels martingales of the Lévy process with scalar product $\langle ., . \rangle_2$ provided by

$$\langle H, J\rangle_2 = \mathbb{E}\left(\left[H, J\right]_1\right) \ , \ H, J \in S_2 \quad , \tag{1.4.7}$$

where $[H, J]$ is the *cross-variation* of H and J.
Nualart and Schoutens [2000] show that the correspondence $x^{i-1} \leftrightarrow Y^{(i)}$ is an *isometry* (note that $x^{i-1} \in S_1$ and $Y^{(i)} \in S_2$). Therefore, an orthogonalization of $\{1, x, x^2, \ldots\}$ in S_1 gives an orthogonalization of $\left\{Y^{(1)}, Y^{(2)}, Y^{(3)}, \ldots\right\}$. The implication of this result is that the orthogonalization of the Teugels martingales used to provide the predictable representation of Lévy processes is intrinsically related to classical orthogonal polynomials. As an example, the orthogonalization of the Teugels martingales of a Gamma process is achieved through the use of *Laguerre polynomials*, that of a negative binomial (or Pascal) process involves *Meixner polynomials,* while in the case of the Brownian motion we use *Hermite polynomials*.

1.5 Upper Bounds in a Lévy Setting Without Nested Simulation

We develop a non-nested Monte Carlo simulation method to derive upper bounds for Bermudan options when asset prices are driven by Lévy processes which extends and generalizes the approach introduced in Belomestny et al. [2009], where the analysis is limited to the Brownian Motion setting. In particular, this methodology takes advantage of the results on the predictable representation property for Lévy processes of Nualart and Schoutens [2000] to derive an estimator for the optimal martingale M^* in (1.3.3).

1.5.1 A martingale estimator for Lévy processes

As an approximation to the continuous monitoring case, we will consider an option that can be exercised at one date from the set $\mathcal{E} = \{T_0, ..., T_\jmath\}$ (Bermudan option), where we will assume that $T_0 = 0$ and define $T := T_\jmath$. Throughout the exposition we will also assume that the pricing measure Q, with its corresponding discounting *numéraire* $\mathcal{N}$, is given and defined on the filtered probability space $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_t^L\right)_{t\geq 0}, Q\right)$, where $\left(\mathcal{F}_t^L\right)_{t\geq 0}$ is the filtration generated by the square integrable Lévy process L with the following Lévy-Khinchin representation (see Cont

and Tankov [2004])

$$\mathbb{E}^{Q}\left[e^{iuL_t}\right] = e^{t\varphi(u)} \tag{1.5.1}$$

where

$$\varphi(u) = iau - \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty}\left(e^{iux} - 1 - iux1_{(|x|<1)}\right)\nu(dx) \ . \tag{1.5.2}$$

According to the Bermudan contract, when exercising at time $T_j \in \mathcal{E}$, the holder of the option receives a discounted payment of the form

$$Z_{T_j} := \Phi\left(T_j, X_{T_j}\right) \tag{1.5.3}$$

where $\Phi(T_j, \cdot)$ is a real valued measurable function and X is a stochastic process adapted to the filtration $\left(\mathcal{F}_t^L\right)_{t\geq 0}$.
Following Belomestny et al. [2009] we rewrite formula (1.3.3) for the Bermudan case:

$$V^{up}(M) := \mathbb{E}^{Q}\left[\max_{0\leq j\leq \mathcal{J}}\left(Z_{T_j} - M_{T_j}\right)\right] \tag{1.5.4}$$

where $M_{T_j}, 0 \leq j \leq \mathcal{J}$ is any martingale with respect to the filtration $\left(\mathcal{F}_{T_j}^L; 0 \leq j \leq \mathcal{J}\right)$ with initial value $M_0 = 0$. As proved in Rogers [2002], $V^{up}(M)$ is an upper bound for the price of the Bermudan option with cash-flow Z_{T_j}. Moreover, the Bermudan price is attained at the martingale part of the Doob decomposition of the discounted price process (Snell envelope). The latter process is denoted by $V_{T_j}^*$.
In practice the exact value of the Snell envelope is not known, however we assume that some approximation V_{T_j} of the Snell envelope is given. If V is a good approximation and it is decomposed in its Doob decomposition

$$V_{T_j} = V_0 + M_{T_j} + U_{T_j} \ , \tag{1.5.5}$$

where the martingale M and the predictable process U start at zero, then we expect $V^{up}(M)$ to be a close upper bound of V_0^*.
U and M can be interpreted as follows:

$$U_{T_{j+1}} - U_{T_j} = \mathbb{E}_{T_j}^{Q}\left[V_{T_{j+1}}\right] - V_{T_j} \tag{1.5.6}$$

$$M_{T_{j+1}} - M_{T_j} = V_{T_{j+1}} - \mathbb{E}^Q_{T_j}\left[V_{T_{j+1}}\right] \tag{1.5.7}$$

In order to construct an estimator for M we can use the predictable representation of Nualart and Schoutens [2000] for square integrable martingales adapted to the filtration generated by L_t. More specifically

$$M_{T_j} = \sum_{p=1}^{\infty} \int_0^{T_j} \phi_s^{(p)} \, \mathrm{d}H_s^{(p)} \ , \ j = 0, ..., \mathcal{J} \tag{1.5.8}$$

where $\phi^{(p)}$, $p = 1, 2, \dots$, are predictable processes and $\left\{H^{(p)}, p = 1, 2, \dots\right\}$ is a set of pairwise strongly orthogonal martingales obtained as linear combinations of Teugels martingales. Our aim is to truncate the infinite sum up to some order q and then to estimate $\phi^{(p)}$.
Our approximate version of the martingale representation (1.5.8) is given by

$$\widetilde{M}_{T_j} = \sum_{p=1}^{q} \int_0^{T_j} \phi_s^{(p)} \, \mathrm{d}H_s^{(p)} \ , \ j = 0, ..., \mathcal{J} \tag{1.5.9}$$

which leads to the following approximation to the change in value of the option between time T_j and T_{j+1}

$$V_{T_{j+1}} - V_{T_j} \approx \sum_{p=1}^{q} \int_{T_j}^{T_{j+1}} \phi_s^{(p)} \, \mathrm{d}H_s^{(p)} + U_{T_{j+1}} - U_{T_j} \ . \tag{1.5.10}$$

Since $\phi^{(p)}$ can only be estimated on a finite number of points, we consider a finer partition of $\mathcal{E}$ given by $\mathcal{P} = \{t_0, \dots, t_\jmath\}$, where $t_0 = 0$ and $t_\jmath = T$. Therefore the discrete time approximation will be

$$V_{T_{j+1}} - V_{T_j} \approx \sum_{p=1}^{q} \left\{ \sum_{t_l \in \mathcal{P}; T_j \le t_l < T_{j+1}} \phi_{t_l}^{(p)} \left(H_{t_{l+1}}^{(p)} - H_{t_l}^{(p)} \right) \right\} + U_{T_{j+1}} - U_{T_j} \ . \tag{1.5.11}$$

If we then multiply both sides by the increment of the k-th orthogonal mar-

tingale and then take conditional expectation with respect to $\mathcal{F}^L_{t_i}$, we obtain

$$\begin{aligned}\mathbb{E}^Q_{t_i}\left[\left(H^{(k)}_{t_{i+1}} - H^{(k)}_{t_i}\right) V_{t_{i+1}}\right] - \mathbb{E}^Q_{t_i}\left[H^{(k)}_{t_{i+1}} - H^{(k)}_{t_i}\right] V_{t_i} \approx \\ \approx \sum_{p=1}^{q} \phi^{(p)}_{t_i} \mathbb{E}^Q_{t_i}\left[\left(H^{(p)}_{t_{i+1}} - H^{(p)}_{t_i}\right)\left(H^{(k)}_{t_{i+1}} - H^{(k)}_{t_i}\right)\right] \\ + \mathbb{E}^Q_{t_i}\left[H^{(k)}_{t_{i+1}} - H^{(k)}_{t_i}\right]\left(U_{t_{i+1}} - U_{t_i}\right)\end{aligned} \tag{1.5.12}$$

then from the martingale property and strong orthogonality of the martingales $H^{(\cdot)}$ the expression simplifies to

$$\phi^{(k)}_{t_i} \approx \frac{1}{(t_{i+1} - t_i) m^{(k)}_2} \mathbb{E}^Q_{t_i}\left[\left(H^{(k)}_{t_{i+1}} - H^{(k)}_{t_i}\right) V_{t_{i+1}}\right] \tag{1.5.13}$$

where $m^{(k)}_2 := \mathbb{E}^Q\left[\left(H^{(k)}_1 - H^{(k)}_0\right)^2\right] = \frac{1}{t_{i+1}-t_i}\mathbb{E}^Q_{t_i}\left[\left(H^{(k)}_{t_{i+1}} - H^{(k)}_{t_i}\right)^2\right]$ is the standardized second moment of the increments of the orthogonal martingale of order k.

We therefore define

$$\widetilde{\phi}^{(k)}_{t_i} := \frac{1}{(\Delta_i)\, m^{(k)}_2} \mathbb{E}^Q_{t_i}\left[\Delta H^{(k)}_{t_i} V_{t_{i+1}}\right] \tag{1.5.14}$$

The corresponding approximation of the martingale M is

$$\widetilde{M}_{t_i} := \sum_{p=1}^{q} \sum_{l=1}^{i} \widetilde{\phi}^{(p)}_{t_l} \Delta H^{(p)}_{t_l} \tag{1.5.15}$$

Assuming, as in most practical applications, an approximation $(\tau_1, \ldots, \tau_\mathcal{J})$ of the optimal stopping policy is available (for example if LSM was used to calculate a lower bound), expression (1.5.14) can be further simplified by using the approximated value of the Snell envelope implied by $(\tau_1, \ldots, \tau_\mathcal{J})$, i.e. $V_{t_i} := \mathbb{E}_{t_i}[Z_{\tau_i}]$. Then we can rewrite (1.5.14) as

$$\widetilde{\phi}^{(k)}_{t_i} := \frac{1}{(\Delta_i)\, m^{(k)}_2} \mathbb{E}^Q_{t_i}\left[\Delta H^{(k)}_{t_i} Z_{\tau_{j+1}}\right], \quad T_j \le t_i \le T_{j+1}, \tag{1.5.16}$$

where we have used the tower property of conditional expectations. Therefore it is not necessary to compute the conditional expectations in the definition of V.

1.5.2 Upper bounds without nested Monte Carlo

We now describe an algorithm based on the construction of the martingales $\widetilde{M}$ that allows us to calculate dual upper bounds without nested Monte Carlo. To this end we suppose that the approximative Snell envelope V_{T_j} is of the form

$$V_{T_j} = u\left(T_j, L_{T_j}\right). \tag{1.5.17}$$

We emphasize that numerical methods to approximate the Snell envelope typically yield approximations of this form. It is then straightforward that the conditional expectations in the definition of $\widetilde{\phi}^{(\cdot)}$ are, in fact, regressions on L_{T_j} . Precisely

$$\widetilde{\phi}_{t_i}^{(k)} = \frac{1}{(\Delta_i)\, m_2^{(k)}} \mathbb{E}_{L_{t_i}}^{Q}\left[\Delta H_{t_i}^{(k)} u\left(T_j, L_{T_j}\right)\right], \quad T_j \leq t_i \leq T_{j+1}. \tag{1.5.18}$$

Next we approximate $\widetilde{\phi}_{t_i}^{(k)}$ by simulation based least squares regression on basis function as was suggested by Longstaff and Schwartz [2001] for lower bounds. To this end we simulate $\widetilde{N}$ independent samples of the Lévy increments ΔL_i, $i = 0, ..., \mathcal{I}$.

Given a vector of basis functions $\varpi^{(k)}\left(t_i, \cdot\right) = \left(\varpi_s^{(k)}\left(t_i, \cdot\right), s = 1, ..., S\right)$, $\widetilde{N}$ independent samples $\left(t_i,\ {}_n\widetilde{L}_{t_i}\right)$, $n = 1, ..., \widetilde{N}$, where ${}_n\widetilde{L}_{t_i} := \sum_{l=0}^{i} \Delta\, {}_nL_{t_l}$, and $\Delta\, {}_n\widetilde{H}_{t_i}^{(k)}$ the increments of the strongly orthogonal martingales constructed on $\Delta\, {}_nL_{t_i}$, we define

$$\hat{\beta}_{t_i}^{(k)} := \underset{\beta \in \mathbb{R}^S}{\arg\min} \left\{ \sum_{n=1}^{\widetilde{N}} \left| \frac{\Delta\, {}_n\widetilde{H}_{t_i}^{(k)}}{(\Delta_i)\, m_2^{(k)}} u\left(T_{j+1},\ {}_n\widetilde{L}_{T_{j+1}}\right) - \varpi^{(k)}\left(t_i,\ {}_n\widetilde{L}_{t_i}\right)\beta \right|^2 \right\}, \tag{1.5.19}$$

with $T_j \leq t_i \leq T_{j+1}$. Then, the corresponding approximative regression mapping for $\widetilde{\phi}_{t_i}^{(k)}$ is defined by

$$\hat{\phi}_{t_i}^{(k)} = \varpi^{(k)}\left(t_i, L_{t_i}\right) \hat{\beta}_{t_i}^{(k)} \tag{1.5.20}$$

After having obtained the functions $\hat{\phi}_{t_i}$ by the above regression procedure, we next construct an approximation of M_{t_i} as follows

$$\hat{M}_{t_i} := \sum_{p=1}^{q} \sum_{l=1}^{i} \hat{\phi}_{t_l}^{(p)} \Delta H_{t_l}^{(p)} \tag{1.5.21}$$

where the increments ΔH are assumed to be independent of the Lévy increments simulated above.
By sampling a new set of N independent trajectories, $\left(T_j,\, {}_nL_{T_j}\right)$, $n = 1, ..., N$ of L an unbiased estimator for $V^{up}\left(\hat{M}\right)$ is obtained by setting:

$$\hat{Y}^{up}\left(\hat{M}\right) := \frac{1}{N} \sum_{n=1}^{N} \max_{0 \leq j \leq \mathcal{J}} \left(\Phi\left(T_j,\, X_{T_j}\left({}_nL_{T_j}\right)\right) - \hat{M}_{T_j} \right) \tag{1.5.22}$$

1.5.3 Variance reduced primal-dual algorithm

As suggested by Belomestny et al. [2009] for the case of standard Itô processes, the martingale $\hat{M}$ obtained from the procedure described above can also be used to reduce the variance in the primal-dual algorithm of Andersen and Broadie [2004].
More specifically, the estimator used by Andersen and Broadie for the upper-bound of a Bermudan option, is given by the following:

$$\begin{aligned} \hat{Y}_{AB}^{up} &:= \frac{1}{N} \sum_{n=1}^{N} \max_{0 \leq j \leq \mathcal{J}} \left(\Phi\left(T_j,\, {}_nX_{T_j}\right) - \sum_{i=1}^{j} \left({}_nV_{T_i} - \frac{1}{K} \sum_{l=1}^{K} \left({}_nV_{T_i,T_{i-1}}^{(l)} \right) \right) \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} \max_{0 \leq j \leq \mathcal{J}} \left(\Phi\left(T_j,\, {}_nX_{T_j}\right) - \; {}_nM_{T_j}^{AB} \right) \end{aligned} \tag{1.5.23}$$

where N is the number of paths used in the Monte Carlo experiment and $\frac{1}{K}\sum_{l=1}^{K}\left({}_nV_{T_i,T_{i-1}}^{(l)}\right)$ is used as an unbiased estimator for $\mathbb{E}_{T_{i-1}}^{Q}\left[V_{T_i}\right]$ (K being the number of sub-paths for the nested simulation and ${}_nV_{T_i,T_{i-1}}^{(l)}$ being a simulated realization of V_{T_i} conditional on the information available at time T_{i-1} on the nth path). From (1.5.7)

it follows that

$$\mathbb{E}^{Q}_{T_i-1}\left[V_{T_i}\right] \approx V_{T_i} - \left(\hat{M}_{T_i} - \hat{M}_{T_{i-1}}\right)$$

therefore

$$\hat{Y}^{up}\left(\hat{M}\right) := \frac{1}{N}\sum_{n=1}^{N}\max_{0\le j\le \mathcal{J}}\Bigg\{ \Phi\left(T_{j,n}\, X_{T_j}\right) - \sum_{n=1}^{j}\left({}_nV_{T_i} - \frac{1}{K}\sum_{l=1}^{K}\left({}_nV^{(l)}_{T_i,T_{i-1}} - \left({}_n\hat{M}^{(l)}_{T_i} - {}_n\hat{M}^{(l)}_{T_{i-1}}\right)\right)\right)\Bigg\} \tag{1.5.24}$$

can be seen as a variance reduced version of the estimator of Andersen and Broadie, with control variate $\hat{M}$.

1.6 Long Step Approach

Here we introduce an easily implementable and computationally efficient variant of the methodology that applies when the value of the underlying financial variables can be expressed using non path-dependent functions of Lévy processes, and requires the simulation of variables only at the exercise dates of the Bermudan product. The condition of non path-dependence implies that the value at time t of the underlyings can be expressed as a function of the time-t value of the Lévy processes driving the model. It should be noted that this property holds for some of the most popular linear and exponential Lévy models that have been proposed in the literature to describe the dynamics of asset prices (e.g.: Black and Scholes [1973], Merton [1976], Madan and Seneta [1990], Carr et al. [2002], Kou [2002], etc.).

Assuming continuity of the value function (cf. Gerhold [2011]), Weierstrass Approximation Theorem can be applied. The theorem states that if $f(x)$ is a continuous function over the closed interval $[a, b]$, then for every $\epsilon > 0$ there exists a polynomial $p(x)$ such that

$$\sup_{x\in[a,b]} |f(x) - p(x)| < \epsilon;$$

in other words, any continuous function on a closed and bounded interval can be uniformly approximated on that interval by a polynomial function to any degree of accuracy. In case the continuation value of the Bermudan contract is a non path-dependent function of a d-dimensional Lévy processes, we can introduce a more parsimonious (and better suited for practical applications) approximation for the change in value of the option between exercise dates compared to (1.5.11). Specifically, let $\boldsymbol{I}$ be an ordered set of d-dimensional indices (i.e. every $\boldsymbol{i}$ in $\boldsymbol{I}$ has the form $\boldsymbol{i} = (i_1, i_2, \ldots, i_d)$ with $i_k \in \mathbb{N}^0$, $k = 1, \ldots, d$) containing $\mathcal{M}$ elements, then we can write

$$V_{T_{j+1}} - V_{T_j} \approx \Psi_{T_j}{}^{\mathsf{T}} \left(G_{T_{j+1}} - G_{T_j}\right) + U_{T_{j+1}} - U_{T_j} \ . \tag{1.6.1}$$

where $\Psi_{T_j} := \left(\psi^{(1)}, \ldots, \psi^{(\mathcal{M})}\right)^{\mathsf{T}}$ is a $\mathcal{M}$-dimensional vector of weights and $G := \left(G^{(\boldsymbol{i})}\right)_{\boldsymbol{i}\in\boldsymbol{I}} = \left(G^{(\mathbf{1})}, \ldots, G^{(\mathcal{M})}\right)^{\mathsf{T}}$ is a vector martingale difference sequence whose elements $G^{(\boldsymbol{i})}$, $\boldsymbol{i} \in \boldsymbol{I}$, are defined as:

$$\begin{aligned} G^{(\boldsymbol{i})}_{T_j} &:= \sum_{1\le k\le j} p^{(\boldsymbol{i})}(\Delta L_{T_k}) - \sum_{1\le k\le j} \mathbb{E}_{T_{k-1}}\left[p^{(\boldsymbol{i})}(\Delta L_{T_k})\right] \\ &= \sum_{1\le k\le j} p^{(\boldsymbol{i})}(\Delta L_{T_k}) - \sum_{1\le k\le j} m^{(\boldsymbol{i})}_k \ , \end{aligned} \tag{1.6.2}$$

with

$$p^{(\boldsymbol{i})}(x) := x_1^{i_1} x_1^{i_2} \ldots x_d^{i_d} \ .$$

Remark 1.6.1. Notice that (1.6.1) has the benefit, compared to (1.5.11), of requiring the estimation of only one set of weight parameters $\left(\psi^{(\boldsymbol{i})}\right)_{\boldsymbol{i}\in\boldsymbol{I}}$ for each step between exercise dates and does not involve the summation of increments over a finer partition. The other significant advantage is that in the Lévy context the 'compensators' $m^{(\boldsymbol{i})}_k$ are constant (so long as there are regular intervals between exercise dates) and can therefore be easily computed by simulation.

1.6.1 The Martingale Estimator

We now want to find a way of estimating the value of Ψ_{T_j} such that the approximation (1.6.1) holds. To do so we define $\Upsilon_{T_j}(\cdot)$ as the second moment, conditional

on the information available at time T_j, of the hedging error between time T_j and T_{j+1}, obtained by applying a set of martingale loadings corresponding to the vector Ψ. More formally

$$\Upsilon_{T_j}(\Psi) := \mathbb{E}^Q_{T_j}\left[\left(\Delta V_{T_{j+1}} - \Delta U_{T_{j+1}} - \Psi^{\mathsf{T}}\Delta G_{T_{j+1}}\right)^2\right] .$$

It follows that our estimate $\widetilde{\Psi}_{T_j}$ is given by

$$\widetilde{\Psi}_{T_j} := \underset{\Psi\in\mathbb{R}^{\mathcal{M}}}{\arg\min}\ \Upsilon_{T_j}(\Psi) . \tag{1.6.3}$$

In order to solve problem (1.6.3) we apply first order conditions; specifically

$$\frac{\partial\Upsilon_{T_j}}{\partial\Psi} = 2\,\mathbb{E}^Q_{T_j}\left[\Delta G_{T_{j+1}}\Delta V_{T_{j+1}} - \Delta G_{T_{j+1}}\Delta U_{T_{j+1}} - \Delta G_{T_{j+1}}\Delta G^{\mathsf{T}}_{T_{j+1}}\Psi\right] = 0 ,$$

from which we can derive the expression for the optimal martingale loadings conditional on the information available at time T_j, i.e.

$$\widetilde{\Psi}_{T_j} = \mathbb{E}^Q_{T_j}\left[\Delta G_{T_{j+1}}\Delta G^{\mathsf{T}}_{T_{j+1}}\right]^{-1}\mathbb{E}^Q_{T_j}\left[\Delta G_{T_{j+1}}\Delta V_{T_{j+1}}\right] ,$$

where we have used the fact that $\mathbb{E}^Q_{T_j}\left[\Delta G_{T_{j+1}}\Delta U_{T_{j+1}}\right] = 0$, which follows from having $\mathbb{E}^Q_{T_j}\left[\Delta G_{T_{j+1}}\right] = 0$ (G is a martingale difference sequence) and from the fact that $\Delta U_{T_{j+1}}$ is measurable with respect to $\mathcal{F}_{T_j}$.

Similarly to the procedure described in Section 1.5.2, the estimation of $\widetilde{\Psi}_{T_j}$ across the discrete-time filtration $\left(\mathcal{F}^L_{T_j}; 0 \le j \le \mathcal{J}-1\right)$ can be performed by simulation using an LSM-like approach. Specifically, we simulate over the exercise times $(T_0,\ldots,T_{\mathcal{J}})$ $\widetilde{N}$ independent paths of the Lévy processes driving the model

$$\widetilde{L}^{(\widetilde{N})} := \left\{ \left({}_n\widetilde{L}_{T_j}\right)_{j=0,\ldots,\mathcal{J}},\ n=1,\ldots,\widetilde{N}\right\}$$

and we use them to calculate the discretized paths of the compensated power-increments vector process

$$\widetilde{G}^{(\widetilde{N})} := \left\{ \left({}_n\widetilde{G}_{T_j}\right)_{j=0,\ldots,\mathcal{J}},\ n=1,\ldots,\widetilde{N}\right\}$$

and the paths of a vector of basis functions

$$\varpi\left(T_j,\cdot\right) := \left(\varpi_1^{(k)}\left(T_j,\cdot\right),\ldots,\varpi_S^{(k)}\left(T_j,\cdot\right)\right)^{\mathsf{T}}.$$

Then, following the LSM approach, we define

$$\hat{\beta}_{T_j} := \underset{\beta\in\mathbb{R}^S}{\arg\min}\left\{\sum_{n=1}^{\widetilde{N}}\left|\left(\widetilde{GG}_{T_{j+1}}\right)^{-1}\Delta({}_n\widetilde{G}_{T_{j+1}})\,u\left(T_{j+1},\,{}_n\widetilde{L}_{T_{j+1}}\right)-\varpi\left(T_j,\,{}_n\widetilde{L}_{T_j}\right)^{\mathsf{T}}\beta\right|^2\right\}, \tag{1.6.4}$$

where

$$\widetilde{GG}_{T_{j+1}} := \frac{1}{\widetilde{N}}\sum_{n=1}^{\widetilde{N}}\Delta({}_n\widetilde{G}_{T_{j+1}})\,\Delta({}_n\widetilde{G}_{T_{j+1}})^{\mathsf{T}}$$

and $u\left(\cdot,\cdot\right)$ is as in (1.5.17), which results in the following estimation of the martingale loadings $\widetilde{\Psi}_{T_j}$

$$\widehat{\Psi}_{T_j} := \varpi\left(T_j,\,L_{T_j}\right)^{\mathsf{T}}\hat{\beta}_{T_j}\,. \tag{1.6.5}$$

Based on this we are now able to obtain our estimate of the martingale $\widetilde{M}$ as

$$\widehat{M}_{T_j} := \sum_{l=1}^{j}(\widehat{\Psi}_{T_{l-1}})^{\mathsf{T}}\Delta G_{T_l}\,, \tag{1.6.6}$$

where we assume that G is simulated independently of the simulation used for estimating $\widehat{\Psi}$.

Remark 1.6.2. Note that at time T_0, because the value of the state variables is the same across all paths, definition (1.6.4) can be simplified as

$$\hat{\beta}_{T_0} := \underset{\beta\in\mathbb{R}^S}{\arg\min}\left\{\sum_{n=1}^{\widetilde{N}}\left|\left(\widetilde{GG}_{T_1}\right)^{-1}\Delta({}_n\widetilde{G}_{T_1})\,u\left(T_1,\,{}_n\widetilde{L}_{T_1}\right)-\beta\right|^2\right\}, \tag{1.6.7}$$

leading to

$$\hat{\beta}_{T_0} = \frac{1}{\widetilde{N}}\sum_{n=1}^{\widetilde{N}}\left(\widetilde{GG}_{T_1}\right)^{-1}\Delta({}_n\widetilde{G}_{T_1})\,u\left(T_1,\,{}_n\widetilde{L}_{T_1}\right) \tag{1.6.8}$$

and

$$\widehat{\Psi}_{T_0} := \hat{\beta}_{T_0}\,.$$

1.6.2 Normalization of Non-orthogonal Martingales

In some cases it may be useful to ortho-normalize the martingale basis $G_t^{(\boldsymbol{i})}$, $\boldsymbol{i} \in \boldsymbol{I}$, with respect to the probability distribution of ΔL in order to improve the stability of the regression procedure (cf. Narula [1979]). Again, given the properties of stationarity and independence of the increments of Lévy processes, this can easily be achieved by estimating the covariance matrix of $G_t^{(\boldsymbol{i})}$, $\boldsymbol{i} \in \boldsymbol{I}$, through an independent simulation. The simplest way to proceed is as follows: let Σ be the covariance of $\Delta G^{(\boldsymbol{i})}$ (which may be available in closed form or estimated, out of sample, by Monte Carlo simulation), then we can obtain the increments of the orthonormalized martingales $\overline{G}_t^{(i)}$, $i = 1, \ldots, N$ (with $N = n(\boldsymbol{I})$) as

$$\Delta \overline{G} = \Delta G \, L^{-1}$$

where $L^\top L = \Sigma$ is the Cholesky factorization of Σ. An alternative way to perform the ortho-normalization of the martingale basis is through the modified Gram-Schmidt algorithm (based on covariance Σ) which, as shown in Rice [1966], provides improved numerical stability at the cost of a slightly more complex implementation.

1.7 Numerical Implementation

In this section we provide details of the algorithms that can be used for the implementation of the methodologies introduced in this chapter and we show some numerical examples.

1.7.1 Pricing Algorithms

Lower bound

Similarly to Andersen and Broadie [2004], our approach is based on the use of LSM for the estimation of lower bounds to the option price. The steps involved in the implementation of the LSM algorithm are described below.

Definition 1.7.1 (Lower bound algorithm - LSM)**.** Steps of the algorithm:

1. Simulate $N_{LSM}^{Est.}$ paths of the underlying variables under the pricing probability measure.

2. Calculate the value of basis functions (e.g. polynomials up to degree Q_{LSM}^{Basis}) along the simulate paths .

3. Estimate the optimal exercise policy $\widetilde{\tau}$ at each possible exercise point by least-square regression with respect to the basis functions calculated at point 2.

4. Apply the estimated policy $\widetilde{\tau}$ to a new set of $N_{LSM}^{Sim.}$ simulated paths of the underlying variables.

5. Calculate the lower bound estimate by averaging the discounted payoffs obtained by applying $\widetilde{\tau}$.

Upper bound

In this chapter we introduced two new methodologies for the calculation of upper bounds to the option price. The general approach, described in section 1.5, for computing upper bounds in a Lévy setting without nested simulation can be implemented through the following algorithm.

Definition 1.7.2 (Upper bound algorithm - predictable representation approach)**.** Steps of the algorithm:

1. Use the modified Gram-Schmidt orthonormalization algorithm (cf. Björck [1967]) to derive the coefficients for the orthonogonalization of the Teugels martingales based on the Lévy measure of the driving processes.

2. Simulate $N_{PR}^{Est.}$ paths of the underlying variables under the pricing measure.

3. Calculate the orthogonalized Teugels martingales basis by applying the orthogonalization coefficients to the compensated power-jump processes obtained from the simulated Lévy processes over the fine time grid $\mathcal{P}$.

4. Calculate value of basis functions along the simulated paths (e.g. polynomials up to degree Q_{PR}^{Basis}) over coarse time grid $\mathcal{E} = \{T_0, ..., T_\jmath\}$ (as suggested

in Belomestny et al. [2009], this can be the set of available exercise dates of the Bermudan product).

5. Estimate the coefficients $\hat{\phi}_{T_j}^{(k)}$, $k = 1, ..., Q_{PR}^{Mtgl}$, of the predictable representation (up to degree Q_{PR}^{Mtgl}) of the martinglale component of the Doob-Meyer decomposition of $V_{T_j}^*$ by least-squares regression with respect to the basis functions calculated at point 4.

6. Calculate discounted payoff Z_t of the option and estimated optimal martingale $\widetilde{M}_t$ at each exercise date on a new set of $N_{PR}^{Sim.}$ simulated paths.

7. Using the dual representation (1.5.4), calculate the approximated upper-bound for the option value at time 0; specifically, the upper bound is obtained by averaging $\max_{0\leq j\leq \mathcal{J}} \left(Z_{T_j} - \widetilde{M}_{T_j} \right)$ over the generated paths.

Finally, the steps of the algorithm for the implementation of the long step approach of section 1.6 are given below.

Definition 1.7.3 (Upper bound algorithm - long step approach)**.** Steps of the algorithm:

1. Simulate $N_{LS}^{Est.}$ paths of the underlying variables.

2. Either analytically or by simulation, estimate the mean value $\widetilde{\mu}$ of the power increments of the driving Lévy processes between exercise dates (up to the chosen degree Q_{LS}^{Mtgl}) .

3. Calculate the martingale basis using the compensated power-increments of the Lévy processes between exercise dates, where the compensators are based on the estimates obtained at point 2. Optionally, the power-increment martingales can be orthogonalized (as described in Section 1.6.2) based on an estimate of their covariance $\widetilde{V}$ (obtained in a similar fashion as $\widetilde{\mu}$).

4. Calculate value of basis functions along the simulated paths (e.g. polynomials up to degree Q_{LS}^{Basis}).

5. At time 0 and at each exercise date (i.e. $T_0, \ldots, T_{\mathcal{J}-1}$), estimate coefficients $\widehat{\Psi}_{T_j}$ (up to the chosen degree Q_{LS}^{Mtgl}) of the approximated optimal martingale

$\widetilde{M}$, based on (1.6.5), by least-square regression with respect to the basis functions calculated at point 4.

6. Calculate instrument discounted payoff Z_t and the estimated optimal martingale $\widetilde{M}_t$ at each exercise date on a new set of $N_{LS}^{Sim.}$ simulated paths.

7. Using the dual representation (1.5.4), calculate the approximated upper-bound for the option value at time 0; specifically, the upper bound is obtained by averaging $\max_{0\leq j\leq \mathcal{J}} \left(Z_{T_j} - \widetilde{M}_{T_j} \right)$ over the generated paths.

Remark 1.7.4 (Choice of polynomial family). The LSM algorithm is not directly affected by the choice of polynomial basis used in the least-squares regression because it relies on the fitted value and not on the degree of correlation among the independent variables. However, if the choice of basis functions leads to a nearly singular matrix, then it is possible that some regression algorithms will give inaccurate numerical results for the estimated conditional expectation function. However, in general, as shown by Areal et al. [2008], most commonly used polynomial families (including combination of monomials) provide almost identical results. Even though the choice of polynomial basis is not relevant in terms of accuracy, it does have an impact in terms of computation time. Therefore Areal et al. [2008] recommend, when time is relevant, the use of the combination of monomials polynomial family which is one of the fastest to compute.

1.7.2 Examples

In our examples we have considered models driven by two types of Lévy processes: Brownian motion and Variance-Gamma (VG); where the first was chosen to allow comparisons of the newly introduced methods with benchmark cases from existing literature and the second (see Madan and Seneta [1990] and Madan and Milne [1991]) was selected for being a well known example of a Lévy process with no continuous martingale component (i.e. pure-jump), thereby allowing to highlight the increased domain of applicability of the proposed approaches compared to Belomestny et al. [2009].

Brownian motion For the Brownian setting we assume that the risk-neutral dynamics of asset prices are defined by the following stochastic differential equation:

$$\mathrm{d}X_t^d = (r-\delta)X_t^d\mathrm{d}t + \sigma X_t^d \mathrm{d}W_t^d \ , \quad d = 1,\dots,D \ ; \tag{1.7.1}$$

where W_t^d, $d = 1,\dots,D$, are independent one-dimensional Wiener processes and r, δ, σ are constants. Stochastic differential equation (1.7.1) is also known as geometric Brownian motion (GBM).

Variance Gamma The pure-jump nature of the VG process means that it can be expressed in terms of its Lévy measure $\nu_{VG}(x)$. Adopting the usual (θ,σ,υ) parametrization, the Lévy measure is defined as

$$\nu_{VG}(x) := \frac{1}{\upsilon|x|}\exp\left(\frac{\theta}{\sigma^2}x - \frac{1}{\sigma}\sqrt{\frac{2}{\upsilon}+\frac{\theta^2}{\sigma^2}}\,|x|\right) \ . \tag{1.7.2}$$

where υ, $\sigma > 0$. Following Madan et al. [1998] we assume asset price dynamics for X, under constant risk-free interest rates r, as given by

$$X_t^d = X_0^d \exp\left((r+\omega)\,t + L_t^{VG,d}\right) \ , \quad d = 1,\dots,D \ ,$$

where

$$\omega = \frac{1}{\upsilon}\ln\left(1-\theta\upsilon-\frac{\sigma^2}{2}\upsilon\right) \ ,$$

and where $L^{VG,d}$ are independent one-dimensional VG processes with parameters (θ,σ,υ).

Bermudan Put

Firstly we consider the case of calculating the price at time $t = 0$ of a Bermudan option with one underlying X. The contract has maturity at time $T = 3$, strike price $K = 100$, payoff function $(K - X_t, 0)^+$ (*put* option) and it gives the holder the right to exercise at times $\mathcal{E} = \{\Delta t, 2\Delta t, ..., T-\Delta t, T\}$, where $\Delta t = \frac{1}{3}$.

GBM We consider the following parameters for the asset price process: $X_0 = 90, 100, 110$, $\sigma = 0.2$, $\delta = 0$, $r = 0.05$. (See Table 1.1)

Table 1.1: Bermudan Put (GBM)

X_0	Low_{LSM} (SE)	Up_{LS} (SE)	$\text{Up}_{A\&B}$ (SE)
90	13.0329 (0.0106)	13.0833 (0.0046)	13.0938 (0.0051)
100	8.5219 (0.0096)	8.5533 (0.0056)	8.5678 (0.0048)
110	5.5421 (0.0084)	5.5868 (0.0068)	5.5809 (0.0041)

VG - 1 We consider the following parameters for the asset price process: $X_0 = 90, 100, 110$, $\sigma = 0.2$, $\theta = -0.3$, $\upsilon = 0.5$ $r = 0.05$. (See Table 1.2)

Table 1.2: Bermudan Put (Variance Gamma)

X_0	Low_{LSM} (SE)	Up_{LS} (SE)	$\text{Up}_{A\&B}$ (SE)
90	17.1481 (0.0170)	17.2872 (0.0086)	17.2875 (0.0090)
100	13.4878 (0.0162)	13.5913 (0.0091)	13.5961 (0.0089)
110	10.7483 (0.0153)	10.8263 (0.0089)	10.8244 (0.0085)

VG - 2 We consider the following parameters for the asset price process: $X_0 = 90, 100, 110$, $\sigma = 0.12$, $\theta = -0.14$, $\upsilon = 0.17$ $r = 0.01$;. (See Table 1.3)

Table 1.3: Bermudan Put (Variance Gamma - 2)

X_0	Low_{LSM} (SE)	Up_{LS} (SE)	$\text{Up}_{A\&B}$ (SE)
90	12.8896 (0.0109)	12.9233 (0.0036)	12.9558 (0.0051)
100	7.8069 (0.0096)	7.8392 (0.0043)	7.8554 (0.0046)
110	4.5581 (0.0076)	4.5973 (0.0042)	4.5975 (0.0037)

The results above have been calculated by applying the algorithms described in section 1.7.1 with the following simulation parameters:

$$N_{LSM}^{Sim.} = 1000000,\ N_{LSM}^{Est.} = 1000000,\ Q_{LSM}^{Basis} = 4\ ;$$

$$N_{LS}^{Sim.} = 10000,\ N_{LS}^{Est.} = 1000000,\ Q_{LS}^{Basis} = 14,\ Q_{LS}^{Mtgl} = 6\ ;$$

$$N_{A\&B}^{Sim.} = 10000,\ N_{A\&B}^{Sub.} = 500\ .$$

where $N_{A\&B}^{Sim.}$ is the number of paths and $N_{A\&B}^{Sub.}$ the number of nested simulations used to calculate the dual estimates based on the algorithm of Andersen and

Broadie [2004] (A&B). Note that the long-step approach allows to obtain close upper bounds to the option price with execution times that are $20\% - 40\%$ those of A&B, while using a simple and generic polynomial basis for the estimation of coefficients of the power-increment martingales. Furthermore, we can see that the non-nested approach has fairly uniform standard errors across moneyness levels while A&B tends to perform better in case of out of the money options. The reason for this is that the accuracy of the estimates from nested simulation will typically improve when the payoff from most nested paths is zero while the polynomial approximation of the martingale part of the Doob decomposition of the discounted price process has no corresponding advantage in case of zero payoffs.

Bermudan Max Call

We now test the case of a *Bermudan Max Call* option with two underlying assets, $X^{(1)}$ and $X^{(2)}$. This is a benchmark case that has been previously analysed in Glasserman [2004], Andersen and Broadie [2004] and Belomestny et al. [2009].

As in the previous example, the contract has maturity at time $T = 3$, strike price $K = 100$ and the holder has the right to exercise the contract at times $\mathcal{E} = \{\Delta t, 2\Delta t, ..., T - \Delta t, T\}$, where $\Delta t = \frac{1}{3}$. However, the payoff function is now given by $\Phi(X_{1,t}, X_{2,t}) = (\max(X_{1,t}, X_{2,t}) - K)^+$.

GBM We consider the following parameters for the asset price process: $X_0^{(1)}, X_0^{(2)} = 90, 100, 110$, $\sigma = 0.2$, $\delta = 0$, $r = 0.05$. Table 1.4 shows benchmark results from the literature while Table 1.5 shows the results of the long step method.

Table 1.4: Bermudan Max Call (GBM) - Benchmark results

	Belomestny et al. [2009]		Schoenmakers et al. [2012]		A&B price interval
X_0	Low (SE)	Up (SE)	Low (SE)	Up (SE)	
90	8.0242 (0.0383)	8.0891 (0.0347)	8.0556 (0.0219)	8.15655 (0.0034)	[8.053, 8.082]
100	13.859 (0.0480)	13.958 (0.0434)	13.885 (0.0276)	14.0293 (0.0044)	[13.892, 13.934]
110	21.330 (0.0556)	21.459 (0.0495)	21.3671 (0.0319)	21.5319 (0.0048)	[21.316, 21.359]

Table 1.5: Bermudan Max Call (GBM)

X_0	Low_{LSM} (SE)	Up_{LS} (SE)	$\text{Up}_{A\&B}$ (SE)
90	8.0398 (0.0116)	8.1239 (0.0301)	8.1297 (0.0258)
100	13.8703 (0.0146)	13.9861 (0.0322)	13.9916 (0.0317)
110	21.3008 (0.0171)	21.4821 (0.0301)	21.4919 (0.0380)

VG We consider the following parameters for the asset price process: $X_0^{(1)}, X_0^{(2)} = 90, 100, 110$, $\sigma = 0.2$, $\theta = -0.3$, $\upsilon = 0.5$ $r = 0.05$. (See Table 1.6)

Table 1.6: Bermudan Max Call(Variance Gamma)

X_0	Low_{LSM} (SE)	Up_{LS} (SE)	$\text{Up}_{A\&B}$ (SE)
90	30.9068 (0.0315)	31.2634 (0.0752)	31.302 (0.0773)
100	41.7849 (0.0369)	42.1789 (0.0834)	42.3079 (0.0897)
110	53.1479 (0.0416)	53.7769 (0.0938)	53.8717 (0.1040)

In this numerical example we have run simulations based on the following simulation parameters:

$$N_{LSM}^{Sim.} = 1000000,\ N_{LSM}^{Est.} = 1000000,\ Q_{LSM}^{Basis} = 4\ ;$$

$$N_{LS}^{Sim.} = 1000,\ N_{LS}^{Est.} = 1000000,\ Q_{LS}^{Basis} = 9,\ Q_{LS}^{Mtgl} = 5\ ;$$

$$N_{A\&B}^{Sim.} = 1000,\ N_{A\&B}^{Sub.} = 200\ ;$$

Again, using a generic polynomial basis we are able to obtain tight upper bounds to the option price. This is in contrast to Belomestny et al. [2009] and Schoenmakers et al. [2012], where the implementation algorithms require the knowledge of the analytical (approximation) formula for the discounted price of the corresponding European options and the use of Malliavin calculus techniques to generate an *ad-hoc* set of basis functions. It should be noted, however, that as the dimensionality of the problem increases, the number of (generic) polynomial basis functions required to produce good approximations of the optimal martingale can quickly grow to a point where the time required to perform the least-square estimation of the coefficients of the power-increment martingales (1.6.4) limits the computational benefits of the approach. For those situations, the choice of basis

functions becomes crucial to obtain tight upper bounds with a sufficiently small basis, therefore one may still consider to apply the methods mentioned in Belomestny et al. [2009] and Schoenmakers et al. [2012] for selecting basis functions specific to the problem at hand.

1.8 Conclusion

In this chapter we presented a novel approach for estimating upper bounds for the value of Bermudan products that does not require resorting to nested simulation and is applicable when the underlying variables are driven by Lévy processes. In particular, by taking advantage of the predictable representation property for Lévy processes introduced in Nualart and Schoutens [2000], we were able to generalise Belomestny et al. [2009] to the Lévy setting, thereby extending the applicability of the approach to a range of models that are better suited for capturing empirical properties of asset returns such as heavy-tailed distributions and jumps. In addition, we introduced a wide time-step variant of the methodology which is easily implementable and computationally efficient, requiring the simulation of variables only at the exercise dates of the Bermudan product, and is applicable when the value of the underlying financial variables can be expressed using non path-dependent functions of Lévy processes (as in the most common linear and exponential Lévy models). Numerical examples show that the methodology, particularly in its long step variant, can be used to obtain tight upper bounds to the price of Bermudan products using a generic procedure that does not involve derivation of *ad-hoc* sets of basis functions specific to the problem at hand.

Part II

Estimation of Covariance with Non-Synchronous Observations

We introduce a new estimator of the covariance of two diffusion processes that are observed only at discrete times in a non-synchronous manner which does not require any modification of the data and is thereby free of bias. Compared to related estimators, such as the one of Hayashi and Yoshida [2005], the proposed estimator achieves significantly higher efficiency (around 20 % reduction in RMSE in case of Poisson sampling) by taking into account the level of overlapping between each observation interval and hence assigning appropriate weights to the cross-product of the associated increments. We show that the new estimator is unbiased and consistent and analyse its efficiency compared to alternatives from the literature.

Chapter 2

Efficient Covariance Estimation of Non-Synchronously Observed Diffusion Processes

2.1 Introduction

The advent of widespread availability of time series of transaction prices (tick-by-tick data) for large classes of financial assets has generated significant interest in the past few years in how to best use this new wealth of data to make better inferences about the various features of data generating processes. Among these, the estimation of covariance, which is by far the most widely used measure of co-dependence between financial variables, has received particular attention. The literature has identified a number of challenges with using tick-by-tick datasets for covariance estimation, namely the fact that transactions occur randomly (and are thereby non-synchronous) and the presence of market micro-structure noise affecting high-frequency returns. The non-synchronicity of the observations, in particular, introduces difficulties in applying standard approaches to covariance estimation. Specifically, the traditional 'Realized Covariance' (RC) estimator, which has been analysed in the high-frequency context in Andersen et al. [2003] and Barndorff-Nielsen and Shephard [2004], requires returns to be available on the same time grid; therefore, before being able to apply RC, a dataset needs

first to be *synchronised* by performing the imputation of values over synchronous time points. The most common ways of doing this are to either select the most recent observation on or before each point on the chosen time grid (known as 'last-tick' interpolation) or to apply linear interpolation to the available observations (see Martens [2004] for a review of the most common approaches). However, as has been shown in Renò [2003], Hayashi and Yoshida [2005] and Zhang [2011], both linear and last-tick interpolation can introduce significant bias in the covariance estimates, especially when the size of the intervals on the synchronous time grid is small relative to the frequency of the actual trades. The non-synchronicity of the data is in fact seen (cf. Zhang [2011]) as a significant component of the so called "Epps effect" (first reported in Epps [1979]), the empirical phenomenon whereby correlations appear to be decreasing when sampling frequency increases. To get around this problem Hayashi and Yoshida [2005] (see also Hayashi and Yoshida [2008] and Hayashi and Kusuoka [2008]) introduced a new unbiased estimator of realized covariance that does not require observations to be available on a regular grid. More specifically, suppose we have discrete non-synchronous observations of two security 'prices' - or 'logarithmic prices', depending on the context - $X^{(1)}_{\Pi_1} := \left(X^{(1)}_{t^{(1)}_i}\right)_{i=1,\dots,N^{(1)}}$ and $X^{(2)}_{\Pi_2} := \left(X^{(2)}_{t^{(2)}_j}\right)_{j=1,\dots,N^{(2)}}$, where $\Pi_1 := \left\{t^{(1)}_1,\dots,t^{(1)}_{N^{(1)}}\right\}$, $\Pi_2 := \left\{t^{(2)}_1,\dots,t^{(2)}_{N^{(2)}}\right\}$ and $t^{(1)}_1 = t^{(2)}_1 = 0$, whose values are assumed to be generated by continuous-time Itô semimartingales. The traditional RC estimator of the realised covariance $V := \langle X^{(1)}, X^{(2)}\rangle_T$ (where $T := t^{(1)}_{N^{(1)}} \wedge t^{(2)}_{N^{(2)}}$) , based on last-tick interpolation and sampling frequency $\Delta \in \mathbb{R}^+$, can be defined as

$$\mathrm{RC}_\Delta\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right) := \sum_{i=1}^{N}\left(\hat{X}^{(1)}_{i\Delta} - \hat{X}^{(1)}_{(i-1)\Delta}\right)\left(\hat{X}^{(2)}_{i\Delta} - \hat{X}^{(2)}_{(i-1)\Delta}\right) \tag{2.1.1}$$

where

$$N := \lfloor T/\Delta \rfloor \,;^{1} \qquad \hat{X}^{(l)}_t := X^{(l)}_{\max\left\{t^{(l)}_i \;\mid\; t^{(l)}_i \in \Pi_l \,;\; t^{(l)}_i \le t\right\}}, \; l = 1, 2;$$

[1] $\lfloor\cdot\rfloor$ is the floor operator, i.e. $\lfloor x \rfloor = \max\{n \in \mathbb{Z} | n \le x\}$

while the Hayashi-Yoshida estimator (HY) is given by

$$\mathrm{HY}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right) := \sum_{i,j}\left(X^{(1)}_{t^{(1)}_i} - X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(2)}_j} - X^{(2)}_{t^{(2)}_{j-1}}\right) 1_{\{(t^{(1)}_i, t^{(1)}_{i-1}]\cap(t^{(2)}_j, t^{(2)}_{j-1}]\neq\varnothing\}} \tag{2.1.2}$$

Even though the HY estimator (2.1.2) has been proven to be consistent and asymptotically normally distributed (cf. Hayashi and Yoshida [2005] and Hayashi and Yoshida [2008]) and is free of the asynchronicity bias affecting RC, it has the drawback of not being efficient in the way it uses the available observations. One reason for this is that the sum in (2.1.2) assigns equal weights to the cross-products of all overlapping increments of $X^{(1)}$ and $X^{(2)}$, irrespective of the amount of overlap between them. To get an intuition of the problem, let $X^{(1)}$ and $X^{(2)}$ be two monthly time series (where each month contains 30 days), $X^{(1)}$ having observations on the first day and $X^{(2)}$ on last day of the every month. From the point of view of the HY estimator, cross-products of increments with 29 out 30 days in common are just as informative about the covariance of the processes as cross-products of increments sharing a single day. More formally, if we assume $X^{(1)}$ and $X^{(2)}$ are observed on equally spaced grids with a constant offset of $h > 0$ from each other (i.e. $t^{(2)}_i = t^{(1)}_i + h\,,\ \ i = 1, \ldots, N^{(2)}$), we have

$$\lim_{h\to 0^+} \mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1+h}\right) = \mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1}\right) \neq \lim_{h\to 0^+} \mathrm{HY}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1+h}\right)$$

in particular, assuming $t^{(1)}_i - t^{(1)}_{i-1} = \Delta$, $i = 1, \ldots, N$, we have

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[\mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1}\right)\right] &= \Delta^2\, n\, \sigma^2_{1,2} + \Delta^2\, n\, \sigma_{1,1}\, \sigma_{2,2}\ ; \\ &\leq \\ \mathbb{V}\mathrm{ar}\left[\lim_{h\to 0^+} \mathrm{HY}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1+h}\right)\right] &= \Delta^2\, n\, \sigma^2_{1,2} + \Delta^2\, (2\,n - 1)\, \sigma_{1,1}\, \sigma_{2,2}\ . \end{aligned} \tag{2.1.3}$$

Proof. See appendix A.1. □

which highlights the inefficiency of HY (note that RC is the mean-variance efficient estimator in the case of synchronous observations [1]).

[1] see Greene [2008]

In this chapter we introduce a new covariance estimator, referred to as the *Linearly Weighted* (LW) estimator, that is an unbiased covariance estimator which generalizes RC to non-synchronous observations. Compared to the HY estimator, it achieves higher efficiency by taking into account the level of overlapping between each observation interval and hence assigning appropriate weights to the cross-product of the associated increments. Specifically, the estimator applies a higher weight when the intervals closely overlap, therefore resulting in a less 'noisy' sample, and lower weight when the overlap is limited. The weights used can be interpreted as being the result of applying traditional sample covariance estimation at infinitely high frequency to the paths obtained from the linear interpolation of the observed values (hence the "linearly weighted" denomination).

Unbiasedness and consistency of the LW estimator is proved for the case of constant convariance and for the case of time-dependent (deterministic) covariance under the assumption of stationary distributed observation times. Simulation results show a reduction of around 20% in RMSE compared to the Hayashi-Yoshida estimator (HY) in the case of Poisson observation times.

During the work on this chapter, Corsi et al. [2012] and Shephard and Xiu [2012] have independently introduced a QMLE approach to covariance estimation with missing observations and i.i.d market micro-structure noise (see also Shumway and Stoffer [1982]) which, by iterating through a Kalman filter and an Expectation-Maximization algorithm (KEM), obtains a covariance estimate that is shown to achieve the non-parametric efficiency bound. However, the statistical efficiency of their approach comes at the cost of a fairly complex implementation and slow computation, particularly in high dimensions.

The LW estimator is somewhat related to KEM, in the sense that LW can be seen as the result of a similar QMLE with EM approach, where the conditional expectations required by the EM algorithm are computed with respect to a subset of the information available, specifically by looking at each time series in isolation. This difference leads to an estimator which is very simple to implement and fast to compute since it is available in closed form, without requiring multiple iterations, while at the same time preserving most of the efficiency gains of KEM.

2.2 Problem setup

In this chapter we are interested in the estimation of the cumulative covariance between two continuous-time diffusion processes, $X^{(1)}$ and $X^{(2)}$, which are being observed at random times and are therefore not guaranteed to be neither synchronous nor equally spaced. The stochastic processes that we are considering are one-dimensional Itô processes which can be described by the following stochastic differential equation

$$\mathrm{d}X_t^{(l)} = \mu_{l,t}\mathrm{d}t + \sigma_{l,t}\mathrm{d}W_{l,t}\,, \qquad X_0^{(l)} = x_l\,, \qquad l = 1, 2\,;$$

with $\langle \mathrm{d}W \rangle_t = \rho_t \mathrm{d}t$, where ρ_t is an unknown function valued in $(-1, 1)$, $x_l > 0$ is a constant, μ_l is a (possibly unknown) predictable locally bounded drift, and $\sigma_l > 0$ is a deterministic and bounded (possibly unknown) function.

2.3 Non-Synchronous Observations and Downward Biases

The RC estimator (2.1.1) has been used extensively for the estimation of covariance from time series data. However, as we mentioned before, the RC estimator assumes that all time series under consideration have a complete set of observations available on the same time-grid $\Pi := \{t_1, \ldots, t_N\}$. Therefore, whenever a time series $X^{(l)}$ does not have value at a grid point t_i, $i = 1, \ldots, N$, the practitioner will need to find a way of filling it with some value $\hat{X}^{(l)}$ before RC can be applied. This issue is particularly relevant when we deal with datasets of transaction prices (or quotes) whose observation times are random and the probability of two observations happening at the same time can be close to zero. The most common ways of producing synchronous samples from non-synchronous datasets have the potential of introducing significant bias in the covariance estimate, which tends to be more severe as the sampling frequency increases (Δ decreases in (2.1.1)). Therefore the practitioner is faced with a trade-off between introducing a higher bias and sub-sampling the dataset, thereby not taking advantage of a large portion of the information available and obtaining more volatile

estimates.

2.3.1 Linear interpolation

As reported in Barucci and Reno [2002], applying the RC estimator to linearly interpolated time series can result in significantly biased estimates. In particular, increasing the sampling frequency we see the variance of the estimator reducing at the cost of an increased downward bias in mean. Furthermore, in the case of linear interpolation this issue applies not only to covariance estimates but also to those of variance (cf. Dacorogna [2001], Kanatani [2004]).

2.3.2 Last-tick interpolation

Given a stochastic process X for which we have observations available at times $\Pi := \{t_1, \ldots, t_N\}$, the value of the process at time t from 'last-tick' interpolation is given by

$$\hat{X}_t := X_{\max\{t_i \mid t_i \in \Pi;\ t_i \leq t\}} \ .$$

Last-tick interpolation is typically preferred to linear interpolation because it does not introduce obvious extraneous bias when estimating quadratic variations of univariate processes (see Dacorogna [2001], Hansen and Lunde [2004], Kanatani [2004]) processes via (2.1.1) with $X^{(1)} = X^{(2)}$. However, Hayashi and Yoshida [2005] and, more generally, Zhang [2011] show that previous-tick interpolation of non-synchronous time series leads to bias in the covariance estimates obtained by using the RC estimator.

2.4 Tick-by-tick estimators

Hayashi and Yoshida [2005] initiated a new line of research in the covariance estimation literature by introducing an unbiased estimator which takes advantage of tick-by-tick observations of non-synchronous samples without requiring any pre-processing or subsampling of the available data. Their estimator has been extensively analysed in different settings by the recent literature and modifications have been introduced to deal with specific issues such as microstructure effects,

lead-lag relationship and rounded time stamps; notable examples can be found in Voev and Lunde [2007],Griffin and Oomen [2011] Palandri [2006], Audrino and Corsi [2008] and Bibinger [2011].

2.4.1 The Hayashi-Yoshida covariance estimator

We now provide a formal definition of the Hayashi-Yoshida covariance estimator (HY). Let $T \in (0,1)$ be an arbitrary terminal time for observing $X^{(l)}$'s. In addition, let $\Pi_1 := \left(t_i^{(1)}\right)_{i=1,2,\ldots,N^{(1)}}$ and $\Pi_2 := \left(t_i^{(2)}\right)_{i=1,2,\ldots,N^{(2)}}$ be increasing sequences of random observation times such that $t_1^{(1)}, t_1^{(2)} \geq 0$ and $t_{N^{(1)}}^{(1)}, t_{N^{(2)}}^{(2)} \leq T$ and let time-intervals between consecutive observations be defined as $I_i^1 := \left(X_{i-1}^{(1)}, X_i^{(1)}\right]$, with $I^1 := \bigcup_{i=1}^{N^{(1)}} \left(I_i^1\right)$, and $I_j^2 := \left(X_{j-1}^{(2)}, X_j^{(2)}\right]$, with $I^2 := \bigcup_{j=1}^{N^{(2)}} \left(I_j^2\right)$. Let also $|\cdot| : \mathcal{B} \to \mathbb{R}$ represent the length of an interval. Then the HY estimator of the realized covariance between $X^{(1)}$ and $X^{(2)}$, over the period $|I^1 \cap I^2| = \left(t_1^{(1)} \vee t_1^{(2)},\ t_{N^{(1)}}^{(1)} \wedge t_{N^{(2)}}^{(2)}\right]$, is given by

Definition 2.4.1 (Hayashi-Yoshida Covariance Estimator)**.**

$$\begin{aligned} \mathrm{HY}\left(X_{\Pi_1}^{(1)}, X_{\Pi_2}^{(2)}\right) &:= \sum_{i,j} \left(X_{t_i^{(1)}}^{(1)} - X_{t_{i-1}^{(1)}}^{(1)}\right)\left(X_{t_j^{(2)}}^{(2)} - X_{t_{j-1}^{(2)}}^{(2)}\right) 1_{\{(t_i^{(1)}, t_{i-1}^{(1)}] \cap (t_j^{(2)}, t_{j-1}^{(2)}] \neq \varnothing\}} \\ &= \sum_{i,j} \Delta X_i^{(1)} \Delta X_j^{(2)} 1_{\{I_i^1 \cap I_j^2 \neq \varnothing\}} \end{aligned}$$

where $\Delta X_i^{(1)} := X_{t_i^{(1)}}^{(1)} - X_{t_{i-1}^{(1)}}^{(1)}$ and $\Delta X_j^{(2)} := X_{t_j^{(2)}}^{(2)} - X_{t_{j-1}^{(2)}}^{(2)}$. Definition 2.4.1 can also be used to derive the HY estimator for the average covariance over the same period. Specifically

Definition 2.4.2 (Hayashi-Yoshida Average Covariance Estimator)**.**

$$\hat{\mathrm{HY}}\left(X_{\Pi_1}^{(1)}, X_{\Pi_2}^{(2)}\right) := \frac{1}{|I^1 \cap I^2|} \sum_{i,j} \Delta X_i^{(1)} \Delta X_j^{(2)} 1_{\{I_i^1 \cap I_j^2 \neq \varnothing\}} .$$

The unbiasedness (in case of zero drift) and consistency of the HY estimator was proved in Hayashi and Yoshida [2005] under the assumption of independence of the observation times of $X^{(1)}$ and $X^{(2)}$. As mentioned in the introduction, one

issue with the estimator is in the relatively inefficient way in which it uses the available information.

2.4.2 Skip-k sampling HY estimator

Griffin and Oomen [2011] show that, when observations are contaminated by market microstructure noise, it may not be optimal to apply the HY estimator by sampling prices at the highest available observation frequency because this can lead to an accumulation of noise that more than offsets the gains from using more data and may result in biased estimates. Under the assumption of Poisson sampling and *i.i.d.* noise, they then propose a "skip-k sampling" HY estimator whereby prices for both assets are sampled every kth observation, i.e. $t_k^{(l)}, t_{2k}^{(l)}, \ldots, t_{\lfloor N_j^{(l)}/k \rfloor k}^{(l)}$, $l = 1, 2$, and identify explicitly the MSE minimizing skip-k sampling frequency as a function of the the noise level.

2.5 The Linearly Weighted Covariance Estimator: Main Result

We propose a new kind of tick-by-tick covariance estimator which aims at improving the efficiency of the HY estimator, while preserving its simplicity and computational performance. We assume that the sampling times $\Pi := (\Pi_1, \Pi_2)$ satisfy the following conditions:

Condition (A)**.**

1. (I_i^1) and (I_j^2) are independent of $X^{(1)}$ and $X^{(2)}$.
2. As $n(\Pi_1 \cup \Pi_2) \to \infty$, $\mathbb{E}\left[\max_i |I_i^1| \vee \max_j |I_j^2|\right] = o(1)$.[1]

Condition (B)**.** [2]

1. Π_1 and Π_2 are such that $(|I_i^1|)_{i=1,2,\ldots}$ and $(|I_j^2|)_{j=1,2,\ldots}$ are strictly stationary sequences of independent $\mathbb{R}^+$-valued random numbers.

[1] $n(\cdot)$ is the cardinality measure

[2] Required only for the case of non-constant covariance

The quantity to be estimated is the (deterministic) covariation of X_1 and X_2, i.e.:

$$\theta := \int_0^T \rho_t \, \sigma_{1,t} \sigma_{1,t} \mathrm{d}t$$

Definition 2.5.1. (Linearly Weighted covariance estimator)**.**
Average covariance:

$$\mathrm{LW}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right) := \frac{1}{\phi} \sum_{i,j} \Delta X^{(1)}_i \Delta X^{(2)}_j w_{i,j} \ ; \tag{2.5.1}$$

Cumulative covariance:

$$\overline{\mathrm{LW}}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right) := \frac{|I^1 \cap I^2|}{\phi} \sum_{i,j} \Delta X^{(1)}_i \Delta X^{(2)}_j w_{i,j} \ ; \tag{2.5.2}$$

where

$$\phi := \sum_{i,j} |I^1_i \cap I^2_j| w_{ij} \ ,$$

$$w_{ij} := \frac{|I^1_i \cap I^2_j|}{|I^1_i||I^2_j|} \ .$$

Remark 2.5.2. The weights used by the proposed estimator can be interpreted as being the result of applying traditional sample covariance estimation at infinitely high frequency to the paths ($\hat{X}_1$ and $\hat{X}_2$) obtained from the linear interpolation of the observed values of X_1 and X_2, and then applying a bias correction $\hat{\phi}$ (as defined below). Consider, for example, calculating sample covariance of increments $\hat{X}_1$ and $\hat{X}_2$ over the interval $(0, T]$, based on increments of length h. Then

we have:

$$
\begin{aligned}
&\lim_{h\to 0}\frac{1}{\hat{\phi}}\sum_{l=1}^{\lfloor T/h\rfloor}\Delta\hat{X}_{1,l}\Delta\hat{X}_{2,l}\\
=&\lim_{h\to 0}\frac{1}{\sum_{l=1}^{\lfloor T/h\rfloor}\frac{h}{|I^1_{i(l)}|}\frac{h}{|I^2_{j(l)}|}\left|I^1_{i(l)}\cap I^2_{j(l)}\right|}\sum_{l=1}^{\lfloor T/h\rfloor}\left(\frac{h}{|I^1_{i(l)}|}\Delta X^{(1)}_{i(l)}\right)\left(\frac{h}{|I^2_{j(l)}|}\Delta X^{(2)}_{j(l)}\right)\\
=&\lim_{h\to 0}\frac{1}{\sum_{l=1}^{\lfloor T/h\rfloor}h\frac{\left|I^1_{i(l)}\cap I^2_{j(l)}\right|}{|I^1_{i(l)}||I^2_{j(l)}|}}\sum_{l=1}^{\lfloor T/h\rfloor}h\left(\frac{1}{|I^1_{i(l)}|}\Delta X^{(1)}_{i(l)}\right)\left(\frac{1}{|I^2_{j(l)}|}\Delta X^{(2)}_{j(l)}\right)\\
=&\lim_{h\to 0}\frac{1}{\sum_{i,j}\left(h\sum_{l}^{\lfloor T/h\rfloor}1_{\{i(l)=i\wedge j(l)=j\}}\right)\frac{\left|I^1_i\cap I^2_j\right|}{|I^1_i||I^2_j|}}\sum_{i,j}\left(h\sum_{l}^{\lfloor T/h\rfloor}1_{\{i(l)=i\wedge j(l)=j\}}\right)\frac{1}{|I^1_i|}\Delta X^{(1)}_i\frac{1}{|I^2_j|}\Delta X^{(2)}_j\\
=&\frac{1}{\sum_{i,j}\left|I^1_i\cap I^2_j\right|\frac{\left|I^1_i\cap I^2_j\right|}{|I^1_i||I^2_j|}}\sum_{i,j}\frac{\left|I^1_i\cap I^2_j\right|}{|I^1_i|\left|I^2_j\right|}\Delta X^{(1)}_i\Delta X^{(2)}_j\\
=&\frac{1}{\phi}\sum_{i,j}\Delta X^{(1)}_i\Delta X^{(2)}_j w_{i,j}\\
=&\mathrm{LW}\left(X^{(1)}_{\Pi_1},X^{(2)}_{\Pi_2}\right)
\end{aligned}
$$

where $i(l):=\{i:\ l\cdot h\in(t_{1,i-1},t_{1,i}]\}$ and $j(l):=\{j:\ l\cdot h\in(t_{2,j-1},t_{2,j}]\}$.

2.6 QML estimation with the EM algorithm

This new estimator can also be analysed in the Quasi-Maximum Likelihood framework of Xiu [2010] and Shephard and Xiu [2012]. The approach is Quasi-Likelihood because it generates a likelihood function by mis-specifying the data generating model, assuming that log-prices are a rotated multivariate Brownian motion, i.e. ignoring the fact that covariance may be a time varying process. Following Shephard and Xiu [2012], we write the quasi log-likelihood under the assumption of a synchronous (though possibly not evenly spaced) dataset of n

observations, where $X := \left(X^{(1)}, \ldots, X^{(d)}\right)$ and Σ is a $d \times d$ covariance matrix.

$$\begin{aligned}\log f\left(x_{1:n}; \Sigma\right) =& c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\left(X_{t_k} - X_{t_{k-1}}\right)'\Sigma^{-1}\left(X_{t_k} - X_{t_{k-1}}\right) \\ =& c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\ \mathrm{tr}\left\{\Sigma^{-1}\left(\Delta X_k \Delta X_k{}'\right)\right\}\end{aligned} \tag{2.6.1}$$

If we now consider non-synchronous datasets, some of the values in (2.6.1) may be missing. To overcome this we can apply the Expectation - Maximization (EM) algorithm of Dempster et al. [1977]. EM involves the application of the following two steps until the convergence of the estimate:

1. (*Expectation*). Calculate the expectation of the log-likelihood function conditional on the available observations and on covariance matrix $\hat{\Sigma}_k$ (where $\hat{\Sigma}_0$ is a guess).

2. (*Maximization*). Find $\hat{\Sigma}_{k+1}$ that maximizes the log-likelihood calculated at step 1.

Let X^{Obs} denote the available (possibly non-synchronous) observations for the multivariate process X. Applying the standard *Expectation* step of the EM algorithm to the log-likelihood (2.6.1) we obtain

$$\begin{aligned}\mathbb{E}\left[\log f\left(x_{1:n}; \Sigma\right)|X^{Obs}; \Sigma\right] =& c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\ \mathbb{E}\left[\left(\Delta X_k\right)'\Sigma^{-1}\Delta X_k|X^{Obs}; \Sigma\right] \\ =& c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\ \mathrm{tr}\left\{\Sigma^{-1}\mathbb{E}\left[\Delta X_k\left(\Delta X_k\right)'|X^{Obs}; \Sigma\right]\right\} \\ =& c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\ \mathrm{tr}\left\{\Sigma^{-1}\left[\Delta\tilde{X}_k\Delta\tilde{X}_k{}'\right.\right. \\ &\left.\left. + \left(\mathbb{E}\left[\Delta X_k\left(\Delta X_k\right)'|X^{Obs}; \Sigma\right] - \Delta\tilde{X}_k\Delta\tilde{X}_k{}'\right)\right]\right\} \\ =& c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\ \mathrm{tr}\left\{\Sigma^{-1}\left(\Delta\tilde{X}_k\Delta\tilde{X}_k{}'\right.\right. \\ &\left.\left. + \mathbb{E}\left[\Delta X_k\left(\Delta X_k\right)' - \Delta\tilde{X}_k\Delta\tilde{X}_k{}'|X^{Obs}; \Sigma\right]\right)\right\}\ ,\end{aligned}$$

where $\Delta\tilde{X}_k = \mathbb{E}\left[\Delta X_k | X^{Obs}; \Sigma\right]$, which can be re-written as

$$\begin{aligned}\log \tilde{f}\left(x_{1:n}; \Sigma\right) := c - \frac{1}{2}\sum_{k=2}^{n} \log|\Sigma| - \frac{1}{2}\sum_{k=2}^{n} \frac{1}{\Delta_k}\ \mathrm{tr}\Big\{\Sigma^{-1}\Big(\Delta\tilde{X}_k \Delta\tilde{X}_k{}' \\ + \mathbb{E}\left[\Delta X_k \Delta X_k{}' - \Delta\tilde{X}_k \Delta\tilde{X}_k{}' | \Pi; \Sigma\right]\Big)\Big\} \\ = \mathbb{E}\left[\log f\left(x_{1:n}; \Sigma\right) | X^{Obs}; \Sigma\right] \ .\end{aligned} \tag{2.6.2}$$

Then, by applying the *Maximization* step, we obtain the following update to the covariance estimate

$$\begin{aligned}\hat{\Sigma} =& \frac{1}{n-1}\sum_{k=2}^{n} \frac{1}{\Delta_k}\left(\Delta\tilde{X}_k \Delta\tilde{X}_k{}' + \mathbb{E}\left[\Delta X_k \left(\Delta X_k\right)' - \Delta\tilde{X}_k \Delta\tilde{X}_k{}' | \Pi; \Sigma\right]\right) \\ =& \frac{1}{n-1}\sum_{k=2}^{n} \frac{1}{\Delta_k}\left\{\Delta\tilde{X}_k \Delta\tilde{X}_k{}' + \mathrm{MSE}\left(\Delta\tilde{X}_k | \Pi; \Sigma\right)\right\} \ .\end{aligned} \tag{2.6.3}$$

Looking at equation (2.6.2) one can see that the function $\log \tilde{f}\left(x_{1:n}; \Sigma\right)$ that we maximize is a modification of the log-likelihood function (2.6.1) where some of the available values have been taken as expectations, thereby resulting in loss of information. In particular, (2.6.2) first uses a predictor to compute $\Delta\tilde{X}_k \Delta\tilde{X}_k{}'$ based on an estimate of the time series values and then replaces the mismatch between $\Delta\tilde{X}_k \Delta\tilde{X}_k{}'$ and the actual (unknown) value of $\Delta X_k \Delta X_k'$ with its expectation. Clearly, the better the quality of the predictor, the lower will be the loss of information when switching from $\log f$ to $\log \tilde{f}$. In the case of the EM algorithm, the chosen predictor (i.e. the conditional expectation with respect to the available observations and the covariance Σ) achieves optimality in mean-squared sense. Motivated by the need to obtain an efficient yet simple and computationally fast covariance estimator, we adopt a simplified predictor of the missing data, one that takes advantage only of the univariate information of each time series and can therefore be calculated independently of the covariance prior. In other words, this simplified predictor is given by the linear interpolation of each time series along the time dimension. The advantage of doing this is that we avoid having to do complex matrix operations at every stage, while, at the same time, taking advantage of what is typically the most significant information for the pre-

diction of a missing value: the value of the stochastic process itself at its closest observations. As we will see below, this approach has the additional advantage of being able to obtain the value of the estimate in a single iteration. More formally, we introduce the following alternative formulation of (2.6.2) based on the 'linear interpolation' predictor:

$$\begin{aligned}\log \hat{f}\left(x_{1:n};\Sigma\right) :=& c-\frac{1}{2}\sum_{k=2}^{n}\log\left|\Sigma\right|-\frac{1}{2}\sum_{k=2}^{n}\frac{1}{\Delta_k}\ \mathrm{tr}\left\{\Sigma^{-1}\left(\Delta\hat{X}_k\Delta\hat{X}_k{}'\right.\right.\\ &\left.\left.+\mathbb{E}\left[\Delta X_k\left(\Delta X_k\right)'-\Delta\hat{X}_k\Delta\hat{X}_k{}'\middle|\Pi;\Sigma\right]\right)\right\}\end{aligned} \tag{2.6.4}$$

where

$$\begin{aligned}\Delta\hat{X}_k &= \left(\mathbb{E}\left[\Delta X_k^{(1)}\middle|X_{1,\cdot}^{Obs};\Sigma\right],\mathbb{E}\left[\Delta X_k^{(2)}\middle|X_{2,\cdot}^{Obs};\Sigma\right],...,\mathbb{E}\left[\Delta X_{d,k}\middle|X_{d,\cdot}^{Obs};\Sigma\right]\right)\\ &= \left(\mathbb{E}\left[\Delta X_k^{(1)}\middle|\Pi_1\right],\mathbb{E}\left[\Delta X_k^{(2)}\middle|\Pi_2\right],...,\mathbb{E}\left[\Delta X_{d,k}\middle|\Pi_d\right]\right)\ .\end{aligned}$$

If follows that the new update equation is given by

$$\hat{\Sigma}=\frac{1}{n-1}\sum_{k=2}^{n}\frac{1}{\Delta_k}\left(\Delta\hat{X}_k\Delta\hat{X}_k{}'+\mathbb{E}\left[\Delta X_k\left(\Delta X_k\right)'-\Delta\hat{X}_k\Delta\hat{X}_k{}'|\Pi;\Sigma\right]\right) \tag{2.6.5}$$

Given the particular nature of the linear interpolation predictor in (2.6.5), each entry of the matrix $\hat{\Sigma}$ is calculated independently therefore we can write the update equation for each pair (l,k), with $l=1,...,d$ and $k=1,...,d$, of dimensions of X. Specifically, the k+1-th iteration of the covariance estimate between $X^{(1)}$

and $X^{(2)}$ will be given by

$$\begin{aligned}
\hat{\sigma}_{1,2}^{k+1} =& \frac{1}{n-1}\sum_{k=2}^{n}\frac{1}{\Delta_k}\Big(\Delta\hat{X}_{1,k}\Delta\hat{X}_{2,k} + \mathbb{E}\left[\Delta X_{1,k}X_{2,k} - \Delta\hat{X}_{1,k}\Delta\hat{X}_{2,k}|\Pi;\Sigma\right]\Big) \\
=& \frac{1}{n-1}\sum_{k=2}^{n}\frac{1}{\Delta_k}\left\{\Delta X_{i(k)}^{Obs,1}\frac{\Delta_k}{|I_{i(k)}^1|}\Delta X_{j(k)}^{Obs,2}\frac{\Delta_k}{|I_{j(k)}^2|} + \hat{\sigma}_{1,2}^k\left(\Delta_k - |I_{i(k)}^1\cap I_{j(k)}^2|\frac{\Delta_k}{|I_{i(k)}^1|}\frac{\Delta_k}{|I_{j(k)}^2|}\right)\right\} \\
=& \frac{1}{n-1}\sum_{k=2}^{n}\left\{\Delta X_{i(k)}^{Obs,1}\Delta X_{j(k)}^{Obs,2}\frac{\Delta_k}{|I_{i(k)}^1||I_{j(k)}^2|} + \hat{\sigma}_{1,2}^k\left(1 - \frac{|I_{i(k)}^1\cap I_{j(k)}^2|}{|I_{i(k)}^1||I_{j(k)}^2|}\Delta_k\right)\right\} \\
=& \frac{1}{n-1}\sum_{i,j}\left\{\Delta X_i^{Obs,1}\Delta X_j^{Obs,2}\frac{|I_i^1\cap I_j^2|}{|I_i^1||I_j^2|} - \hat{\sigma}_{1,2}^k\frac{|I_i^1\cap I_j^2|}{|I_i^1||I_j^2|}|I_i^1\cap I_j^2|\right\} + \hat{\sigma}_{1,2}^k \\
=& \frac{1}{n-1}\sum_{i,j}\left\{\Delta X_i^{Obs,1}\Delta X_j^{Obs,2}w_{i,j} - \hat{\sigma}_{1,2}^k|I_i^1\cap I_j^2|w_{ij}\right\} + \hat{\sigma}_{1,2}^k \\
=& \frac{1}{n-1}\left\{\left(\sum_{i,j}\Delta X_i^{Obs,1}\Delta X_j^{Obs,2}w_{i,j}\right) - \hat{\sigma}_{1,2}^k\phi\right\} + \hat{\sigma}_{1,2}^k \quad ,
\end{aligned}$$

where $i(\cdot)$ and $j(\cdot)$ are functions converting the indices of the complete time series into the indices of the observed values and $w_{i,j}$ and ϕ are as defined in (2.5.1). If we now assume convergence of the algorithm we can set $\hat{\sigma}_{1,2}^{k+1} = \hat{\sigma}_{1,2}^k = \hat{\sigma}_{1,2}$. Therefore we have

$$\begin{aligned}
\frac{1}{n-1}\left\{\sum_{i,j}\Delta X_i^{Obs,1}\Delta X_j^{Obs,2}\,w_{i,j} - \hat{\sigma}_{1,2}\,\phi\right\} &= 0 \\
\sum_{i,j}\Delta X_i^{Obs,1}\Delta X_j^{Obs,2}\,w_{i,j} - \hat{\sigma}_{1,2}\,\phi &= 0
\end{aligned}$$

and thus

$$\hat{\sigma}_{1,2} = \frac{1}{\phi}\sum_{i,j}\Delta X_i^{Obs,1}\Delta X_j^{Obs,2}\,w_{i,j} = \mathrm{LW}\left(X_{\Pi_1}^{(1)}, X_{\Pi_2}^{(2)}\right) \quad ,$$

which is the LW covariance estimator defined in (2.5.1).

2.7 Unbiasedness and Consistency with Non-Constant Covariance

In this section we prove unbiasedness and consistency of the newly introduced LW estimator.

Theorem 2.7.1. *Suppose conditions A. and B. hold.*

1. *If* $\sup_{0\leq t\leq T}|\mu_{k,t}| \in L^4$, $k = 1,2$, *then* $\mathrm{LW}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right) \to \theta$ *in* L^2 *as* $n(\Pi_1 \cup \Pi_2) \to \infty$.

Remark 2.7.2. As is seen in the proof below, if $\mu^l_t = 0$, $0 \leq t \leq T$, then $\mathrm{LW}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right)$ is unbiased.

2.7.1 Unbiasedness of the LW Estimator

First we assume $\mu^l_t = 0$, $0 \leq t \leq T$ for the time being. We will show that $\mathrm{LW}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right) \to \theta$ in L^2 as $n(\Pi_1 \cup \Pi_2) \to \infty$. For simplicity of notation in the rest of the proof we will refer to $\mathrm{LW}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_2}\right)$ such that $n(\Pi_1 \cup \Pi_2) = n$ as LW_n; also, without loss of generality, we assume $T = 1$. For each measurable set I on $[0,\infty)$, we define (signed) measures by

$$c(I) := \int_I \rho_t\, \sigma_{1,t}\sigma_{1,t}\mathrm{d}t\ ,$$

$$c^k(I) := \int_I (\sigma_{k,t})^2\, \mathrm{d}t\ , \quad k = 1,2$$

Moreover, for each measurable set I on $[0,\infty)$, we define

$$\Delta X^k(I) := \int_0^T 1_I(t)\sigma_{k,t}\mathrm{d}W_{k,t}\ , \quad k = 1,2.$$

In addition we introduce the shorthand notation $\Delta X^{(1)}_i := \Delta X_1(I^1_i)$ and $\Delta X^{(2)}_j := \Delta X_2(I^2_j)$.

In order to prove the unbiasedness of the estimator we need to prove that

$$\mathbb{E}\left[\frac{T}{\phi}\sum_{i,j}\Delta X_i^{(1)}\Delta X_j^{(2)}w_{i,j}\right] = \int_0^T \rho_t\,\sigma_{1,t}\sigma_{1,t}\mathrm{d}t = c\left(I^1\cap I^2\right) = \theta$$

We proceed as follows: let $\tilde{w}:(\Omega,\mathbb{R}^+)\to\mathbb{R}^+$ be defined as $\tilde{w}_t := \sum_{i,j} w_{ij}1_{\left\{t\in I_i^1\cap I_j^2\right\}}$. Then

$$\begin{aligned}
\mathbb{E}\left\{\frac{T}{\phi}\sum_{i,j}\Delta X_i^{(1)}\Delta X_j^{(2)}w_{i,j}\right\} &= \mathbb{E}\left\{\frac{T}{\phi}\sum_{i,j}\mathbb{E}\left[\Delta X_i^{(1)}\Delta X_j^{(2)}\middle|\,\Pi\right]w_{i,j}\right\}\\
&= \mathbb{E}\left\{\frac{T}{\phi}\sum_{i,j}c(I_i^1\cap I_j^2)\,w_{i,j}\right\}\\
&= \mathbb{E}\left\{\frac{T}{\int_0^T\tilde{w}_s\,\mathrm{d}s}\sum_{i,j}\left(\int_{I_i^1\cap I_j^2}\rho_t\,\sigma_{1,t}\sigma_{1,t}\mathrm{d}t\right)w_{i,j}\right\}\\
&= T\int_0^T(\rho_t\,\sigma_{1,t}\sigma_{1,t})\,\mathbb{E}\left[\frac{\tilde{w}_t}{\int_0^T\tilde{w}_s\mathrm{d}s}\right]\mathrm{d}t\\
&= T\frac{1}{T}\int_0^T\rho_t\,\sigma_{1,t}\sigma_{1,t}\mathrm{d}t\\
&= \int_0^T\rho_t\,\sigma_{1,t}\sigma_{1,t}\mathrm{d}t
\end{aligned}$$

Where we have used the independence of the increments and the fact that, following from condition B, $\tilde{w}_t$'s are identically distributed for every t.

2.7.2 Consistency of the LW Estimator

We claim that $\mathbb{E}\left[\mathrm{LW}_n^2\right] = \theta^2 + o(1)$ so that $\mathrm{LW}_n\to\theta$ in L^2 as $n\to\infty$. To this end note that

$$\mathbb{E}\left[\mathrm{LW}_n^2\right] = \mathbb{E}\left[\frac{1}{\phi^2}\sum_{i,j,i',j'}\mathbb{E}\left\{\Delta X_i^{(1)}\Delta X_j^{(2)}\Delta X_{i'}^{(1)}\Delta X_{j'}^{(2)}\middle|\,\Pi\right\}w_{ij}w_{i'j'}\right]$$

and decompose the inside summation into four components, specifically

$$\sum_{i,j,i',j'} = \sum_{\substack{i,j,i',j':\\ i'=i,j'=j}} + \sum_{\substack{i,j,i',j':\\ i'=i,j'\neq j}} + \sum_{\substack{i,j,i',j':\\ i'\neq i,j'=j}} + \sum_{\substack{i,j,i',j':\\ i'\neq i,j'\neq j}} =: D_1 + D_2 + D_3 + D_4 \qquad (2.7.1)$$

In the following four subsections we will calculate the expectations that appear in each component.

Case 1 - D_1

The D_1 component in 2.7.1 is given by

$$D_1 = \frac{1}{\phi^2} \sum_{i,j} \mathbb{E}\left\{ \left(\Delta X_i^{(1)}\right)^2 \left(\Delta X_j^{(2)}\right)^2 \middle| \Pi \right\} w_{ij}^2 \ .$$

In order to make the derivation more compact, we introduce the following notation:

$$\begin{aligned} I_{i\setminus j}^1 &:= I_i^1 \setminus I_j^2 \ , \\ \Delta X_{i\cap j}^{(1)} &:= X_{u(I_i^1 \cap I_j^2)}^{(1)} - X_{d(I_i^1 \cap I_j^2)}^{(1)} \ , \\ \Delta X_{i\setminus j}^{(1)} &:= \left(X_{u(I_i^1)}^{(1)} - X_{d(I_i^1)}^{(1)}\right) - \left(X_{u(I_i^1 \cap I_j^2)}^{(1)} - X_{d(I_i^1 \cap I_j^2)}^{(1)}\right) \ ; \end{aligned}$$

and

$$\begin{aligned} I_{j\setminus i}^2 &:= I_j^2 \setminus I_i^1 \ , \\ \Delta X_{i\cap j}^{(2)} &:= X_{u(I_i^1 \cap I_j^2)}^{(2)} - X_{d(I_i^1 \cap I_j^2)}^{(2)} \ , \\ \Delta X_{j\setminus i}^{(2)} &:= \left(X_{u(I_j^2)}^{(2)} - X_{d(I_j^2)}^{(2)}\right) - \left(X_{u(I_i^1 \cap I_j^2)}^{(2)} - X_{d(I_i^1 \cap I_j^2)}^{(2)}\right) \ ; \end{aligned}$$

where $u(I) := \sup\{t : t \in I\}$ and $d(I) := \inf\{t : t \in I\}$. Then

$$\begin{aligned}
\mathbb{E}\left\{\left(\Delta X_i^{(1)}\right)^2\left(\Delta X_j^{(2)}\right)^2\middle|\Pi\right\} &= \mathbb{E}\left\{\left(\Delta X_{i\setminus j}^{(1)}+\Delta X_{i\cap j}^{(1)}\right)^2\left(\Delta X_{j\setminus i}^{(2)}+\Delta X_{i\cap j}^{(2)}\right)^2\middle|\Pi\right\} \\
&= \mathbb{E}\left\{\left(\Delta X_{i\setminus j}^{(1)}\right)^2\left(\Delta X_{i\cap j}^{(2)}\right)^2\middle|\Pi\right\}+\mathbb{E}\left\{\left(\Delta X_{i\cap j}^{(1)}\right)^2\left(\Delta X_{i\cap j}^{(2)}\right)^2\middle|\Pi\right\} \\
&\quad \mathbb{E}\left\{\left(\Delta X_{i\cap j}^{(1)}\right)^2\left(\Delta X_{j\setminus i}^{(2)}\right)^2\middle|\Pi\right\}+\mathbb{E}\left\{\left(\Delta X_{i\setminus j}^{(1)}\right)^2\left(\Delta X_{j\setminus i}^{(2)}\right)^2\middle|\Pi\right\} \\
&= c_1\left(I_{i\setminus j}^1\right)\, c_2\left(I_i^1\cap I_j^2\right)+2\left[c\left(I_i^1\cap I_j^2\right)\right]^2+c_1\left(I_i^1\cap I_j^2\right)\, c_2\left(I_i^1\cap I_j^2\right) \\
&\quad c_1\left(I_i^1\cap I_j^2\right)\, c_2\left(I_{j\setminus i}^2\right)+c_1\left(I_{i\setminus j}^1\right)\, c_2\left(I_{j\setminus i}^2\right) \\
&= c_1\left(I_i^1\right)c_2\left(I_j^2\right)+2c\left(I_i^1\cap I_j^2\right)\ ,
\end{aligned}$$

where we have used the independence of the increments and the fact that $c^1(I_{i\setminus j}^1) = c_1(I_i^1) - c_1(I_i^1 \cap I_j^2)$ and $c_2(I_{j\setminus i}^2) = c_2(I_j^2) - c_2(I_i^1 \cap I_j^2)$. Therefore we have

$$D_1 = \frac{1}{\phi^2}\sum_{i,j} c_1\left(I_i^1\right)c_2\left(I_j^2\right)w_{ij}^2+\frac{2}{\phi^2}\sum_{i,j} c\left(I_i^1\cap I_j^2\right)^2 w_{ij}^2\ .$$

First term of D_1. Noting that σ^k, $k = 1, 2$, are bounded, we have

$$\begin{aligned}
\frac{1}{\phi^2}\sum_{i,j} c_1\left(I_i^1\right)c_2\left(I_j^2\right)w_{ij}^2 &= \frac{1}{\phi^2}\sum_{i,j}\left(\int_{I_i^1}\left(\sigma_{1,t}\right)^2\mathrm{d}t\right)\left(\int_{I_j^2}\left(\sigma_{2,t}\right)^2\mathrm{d}t\right)w_{ij}^2 \\
&\leq \sup_{0\leq t\leq T}\left(\sigma_{1,t}\right)^2\sup_{0\leq t\leq T}\left(\sigma_{2,t}\right)^2\frac{1}{\phi^2}\sum_{i,j}|I_i^1||I_j^2|w_{ij}^2
\end{aligned}$$

We claim that

$$\mathbb{E}\left[\frac{1}{\phi^2}\sum_{i,j}|I_i^1||I_j^2|w_{ij}^2\right] = o\left(1\right)\ .$$

To this end we decompose

$$\begin{aligned}\frac{1}{\phi^2}\sum_{i,j}|I_i^1||I_j^2|w_{ij}^2 &= \frac{1}{\phi^2}\sum_{i,j}|I_i^1||I_j^2|\left(\frac{|I_i^1\cap I_j^2|}{|I_i^1||I_j^2|}\right)^2\\ &= \frac{1}{\phi^2}\sum_{i,j}w_{ij}|I_i^1\cap I_j^2|\\ &= \frac{1}{\phi^2}\,\phi\\ &= \frac{1}{\phi}\,.\end{aligned}$$

Therefore, in order for $\mathbb{E}\left[\frac{1}{\phi^2}\sum_{i,j}|I_i^1||I_j^2|w_{ij}^2\right]$ to converge to zero we must have $\mathbb{E}[\phi]\to\infty$ as $M\to 0$.

Let's assume now $|I_i^1|,|I_j^2|\le M$, $\forall i,j$. Notice that

$$\begin{aligned}A_h := \sum_k\frac{|I_h^1\cap I_k^2|^2}{|I_h^1||I_k^2|} &\ge \frac{\{\alpha_h|I_h^1|\}^2}{|I_h^1|M}+\frac{(1-\alpha_h-\beta_h)\,|I_h^1|}{|I_h^1|}+\frac{\{\beta_h|I_h^1|\}^2}{|I_h^1|M}\\ &= \frac{|I_h^1|}{M}\left(\alpha_h^2+\beta_h^2\right)+1-\alpha_h-\beta_h =: f_h\left(\alpha_h,\beta_h\right),\end{aligned}$$

where $0\le\alpha_h\le 1$ represents the fraction of $|I_h^1|$ that overlaps with I_m^2, where $I_m^2\cap I_{h-1}^1\neq\emptyset$, and $0\le\beta_h\le 1$ represents the fraction of $|I_h^1|$ that overlaps with I_n^2, where $I_n^2\cap I_{h+1}^1\neq\emptyset$ and $I_n^2\neq I_m^2$, and where $\alpha_h+\beta_h\le 1$. The function $f_h()$ is minimised for $\hat\alpha=\hat\beta=\frac{1}{2}\frac{M}{|I_h^1|}$, which, given the condition $\alpha+\beta\le 1$, reduces to $\hat\alpha=\hat\beta=\frac{1}{2}$.

Therefore we have

$$\begin{aligned}\phi &= \sum_h A_h \ge \sum_h f_h\left(\alpha_h,\beta_h\right)\\ &\ge \sum_h f_h\left(\hat\alpha,\hat\beta\right) = \sum_h\frac{1}{4}\frac{|I_h^1|}{M} = \frac{1}{4M}\sum_h|I_h^1| = \frac{1}{4}\frac{|I^1|}{M}\,.\end{aligned}\tag{2.7.2}$$

From this it follows that, as $M\to 0$, $\mathbb{E}[\phi]\to\infty$.

Second Term of D_1.

$$\frac{2}{\phi^2}\sum_{i,j} c\left(I_i^1 \cap I_j^2\right)^2 w_{ij}^2 = \frac{1}{\phi^2}\sum_{i,j}\left(\int_{I_i^1\cap I_j^2} \sigma_{1,t}\sigma_{2,t}\rho_t \mathrm{d}t\right)^2 w_{ij}^2$$
$$\leq \left(\sup_{0\leq t\leq T}\sigma_{1,t}\sup_{0\leq t\leq T}\sigma_{2,t}\sup_{0\leq t\leq T}|\rho_t|\right)^2 \frac{1}{\phi^2}\sum_{i,j}|I_i^1\cap I_j^2|^2 w_{ij}^2$$

This time we claim that

$$\mathbb{E}\left[\frac{1}{\phi^2}\sum_{i,j}|I_i^1\cap I_j^2|^2 w_{ij}^2\right] = o(1)$$

then

$$\frac{1}{\phi^2}\sum_{i,j}|I_i^1\cap I_j^2|^2 w_{ij}^2 = \frac{\sum_{i,j}\left(|I_i^1\cap I_j^2|w_{ij}\right)^2}{\left(\sum_{i,j}|I_i^1\cap I_j^2|w_{ij}\right)^2}$$

which converges to zero because $0 \leq |I_i^1 \cap I_j^2| w_{ij} \leq 1$ and because, as we have shown above, as M goes to zero, $\mathbb{E}[\phi]$ goes to infinity.

It follows that $\mathbb{E}\left[D_1\right] = o(1)$.

Case 2 - D_2

$$D_2 = \frac{1}{\phi^2}\sum_{i,j,j':j\neq j'}\mathbb{E}\left\{\left(\Delta X_i^{(1)}\right)^2 \Delta X_j^{(2)}\Delta X_{j'}^{(2)}\middle|\,\Pi\right\} w_{ij}w_{ij'}$$

Let

$$\Delta X_{i\setminus(j,j')}^{(1)} := \left(X_{u(I_i^1)}^{(1)} - X_{d(I_i^1)}^{(1)}\right) - \left(X_{u(I_i^1\cap[I_j^2\cup I_{j'}^2])}^{(1)} - X_{d(I_i^1\cap[I_j^2\cup I_{j'}^2])}^{(1)}\right) .$$

Then, using the independence of increments we have

$$\begin{aligned}
\mathbb{E}\left\{ \left(\Delta X_i^{(1)}\right)^2 \Delta X_j^{(2)} \Delta X_{j'}^{(2)} \middle| \Pi \right\} &= \mathbb{E}\left\{ \left(\Delta X_i^{(1)}\right)^2 \Delta X_{i\cap j}^{(2)} \Delta X_{i\cap j'}^{(2)} \middle| \Pi \right\} \\
&= \mathbb{E}\left\{ \left(\Delta X_{i\cap j}^{(1)} + \Delta X_{i\cap j'}^{(1)} + \Delta X_{i\setminus(j,j')}^{(1)}\right)^2 \Delta X_{i\cap j}^{(2)} \Delta X_{i\cap j'}^{(2)} \middle| \Pi \right\} \\
&= 2\,\mathbb{E}\left\{ \Delta X_{i\cap j}^{(1)} \Delta X_{i\cap j}^{(2)} \middle| \Pi \right\} \mathbb{E}\left\{ \Delta X_{i\cap j'}^{(1)} \Delta X_{i\cap j'}^{(2)} \middle| \Pi \right\} \\
&= 2\, c\left(I_i^1 \cap I_j^2\right) c\left(I_i^1 \cap I_{j'}^2\right)
\end{aligned}$$

Hence,

$$\begin{aligned}
D_2 &= \frac{2}{\phi^2} \sum_{i,j,j':j\neq j'} c\left(I_i^1 \cap I_j^2\right) c\left(I_i^1 \cap I_{j'}^2\right) w_{ij} w_{ij'} \\
&= \frac{2}{\phi^2} \sum_i \left\{ \sum_j c\left(I_i^1 \cap I_j^2\right) w_{ij} \left(\sum_{j'} c\left(I_i^1 \cap I_{j'}^2\right) w_{ij'} - c\left(I_i^1 \cap I_j^2\right) w_{ij} \right) \right\} \\
&= \frac{2}{\phi^2} \sum_i \left(\sum_j c\left(I_i^1 \cap I_j^2\right) w_{ij} \right)^2 - \frac{2}{\phi^2} \sum_{i,j} c\left(I_i^1 \cap I_j^2\right)^2 w_{ij}^2
\end{aligned}$$

First term of D_2.

$$\begin{aligned}
\frac{2}{\phi^2} \sum_i \left(\sum_j c\left(I_i^1 \cap I_j^2\right) w_{ij} \right)^2 &= \frac{2}{\phi^2} \sum_i \left(\sum_j c\left(I_i^1 \cap I_j^2\right) w_{ij} \right)^2 \\
&\leq \left(\sup_{0\leq t\leq T} \sigma_{1,t} \sup_{0\leq t\leq T} \sigma_{2,t} \sup_{0\leq t\leq T} |\rho_t| \right)^2 \frac{2}{\phi^2} \sum_i \left(\sum_j |I_i^1 \cap I_j^2| w_{ij} \right)^2
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\phi^2} \sum_i \left(\sum_j |I_i^1 \cap I_j^2| w_{ij} \right)^2 &= \left(\sum_{i,j} \frac{|I_i^1 \cap I_j^2|^2}{|I_i^1||I_j^2|} \right)^{-2} \sum_i \left(\sum_j \frac{|I_i^1 \cap I_j^2|^2}{|I_i^1||I_j^2|} \right)^2 \\
&= \frac{\sum_i (A_i)^2}{\left(\sum_i A_i\right)^2}
\end{aligned}$$

From the fact that $\sum_j \frac{|I_i^1 \cap I_j^2|}{|I_i^1|} = 1$ and $0 \leq \frac{|I_i^1 \cap I_j^2|}{|I_j^2|} \leq 1, \forall j$, it follows that $A_i \leq 1$. Specifically:

$$A_i := \sum_j \frac{|I_i^1 \cap I_j^2|^2}{|I_i^1||I_j^2|} = \left(\alpha_i \frac{|I_i^1 \cap I_{d(i)}^2|}{|I_{d(i)}^2|} + \beta_i \frac{|I_i^1 \cap I_{u(i)}^2|}{|I_{u(i)}^2|} \right) + 1 - \alpha_i - \beta_i \leq 1,$$

where $d(i)$ is such that $I_{d(i)}^2 \cap I_{h-1}^1 \neq \emptyset$ and where $u(i)$ is such that $I_{u(i)}^2 \cap I_{h+1}^1 \neq \emptyset$. Given that $0 \leq A_i \leq 1$ and that, as we showed before, $\phi = \sum_i A_i \to \infty$, it follows that the first term of D_2 is $o(1)$.

Second term of D_2. The second term of D_2 is $o(1)$ as we have already shown in case 1.

Case 3 - D_3

The same argument as in Case 2 applies by symmetry to obtain $\mathbb{E}[D_3] = o(1)$.

Case 4 - D_4

Following Hayashi and Yoshida [2005], we have

$$\begin{aligned} D_4 &= \frac{1}{\phi^2} \sum_{i,i',j,j':i\neq i',j\neq j'} \mathbb{E}\left\{ \Delta X_i^{(1)} \Delta X_j^{(2)} \Delta X_{i'}^{(1)} \Delta X_{j'}^{(2)} \middle| \Pi \right\} w_{ij} w_{i'j'} \\ &= \frac{1}{\phi^2} \sum_{i,i',j,j':i\neq i',j\neq j'} c\left(I_i^1 \cap I_j^2\right) c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{ij} w_{i'j'} \\ &= \frac{1}{\phi^2} \sum_{ij} c\left(I_i^1 \cap I_j^2\right) w_{ij} \left(\sum_{i',j':i\neq i',j\neq j'} c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{i'j'} \right) \end{aligned}$$

Notice that for fixed i and j the following holds

$$\begin{aligned}\sum_{i',j':i\neq i',j\neq j'} c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{i'j'} &= \sum_{i',j'} c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{i'j'} - c\left(I_i^1 \cap I_j^2\right) w_{ij} \\ &\quad - \sum_{j':j'\neq j} c\left(I_i^1 \cap I_{j'}^2\right) w_{ij'} - \sum_{i':i'\neq i} c\left(I_{i'}^1 \cap I_j^2\right) w_{i'j} \\ &= \sum_{i',j'} c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{i'j'} + c\left(I_i^1 \cap I_j^2\right) w_{ij} \\ &\quad - \sum_{j'} c\left(I_i^1 \cap I_{j'}^2\right) w_{ij'} - \sum_{i'} c\left(I_{i'}^1 \cap I_j^2\right) w_{i'j} \ .\end{aligned}$$

Therefore we have

$$\begin{aligned}D_4 &= \frac{1}{\phi^2} \sum_{i,i',j,j':i\neq i',j\neq j'} c\left(I_i^1 \cap I_j^2\right) c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{ij} w_{i'j'} \\ &= \frac{1}{\phi^2} \sum_{i,j} c\left(I_i^1 \cap I_j^2\right) w_{ij} \Bigg(\sum_{i',j'} c\left(I_{i'}^1 \cap I_{j'}^2\right) w_{i'j'} + c\left(I_i^1 \cap I_j^2\right) w_{ij} \\ &\quad - \sum_{j'} c\left(I_i^1 \cap I_{j'}^2\right) w_{ij'} - \sum_{i'} c\left(I_{i'}^1 \cap I_j^2\right) w_{i'j} \Bigg) \\ &= \frac{1}{\phi^2} \left(\sum_{i,j} c\left(I_i^1 \cap I_j^2\right) w_{ij} \right)^2 + \frac{1}{\phi^2} \sum_{i',j'} c\left(I_{i'}^1 \cap I_{j'}^2\right)^2 w_{i'j'}^2 \\ &\quad - \frac{1}{\phi^2} \sum_{i,j,j'} c\left(I_i^1 \cap I_j^2\right) c\left(I_i^1 \cap I_{j'}^2\right) w_{ij} w_{ij'} - \frac{1}{\phi^2} \sum_{i,j,i'} c\left(I_i^1 \cap I_j^2\right) c\left(I_{i'}^1 \cap I_j^2\right) w_{ij} w_{i'j} \ .\end{aligned}$$

We have already shown that the second, third and fourth elements of the summation are $o(1)$. As for the first, we want to show that

$$\lim_{M\to 0} \mathbb{E}\left[\left(\frac{1}{\phi} \sum_{i,j} c\left(I_i^1 \cap I_j^2\right) w_{ij} - \theta \right)^2 \right] = 0$$

For this proof we are going to use some additional notation/definitions:

- $f : X \to Y$ is such that $f_t = \rho_t \, \sigma_{1,t} \sigma_{2,t}$, $\forall t \in [0,1]$, with $X = (0,1]$ and $Y = \mathbb{R}$.

- $\Pi^Y := (Y_k)_{k=1,2,...}$ is a partition of Y, where $Y_k := (y_{k-1}, y_k]$, such that $|Y_k| \leq N$, $\forall k \in \mathbb{N}$
- $\Pi^X := (X_k)_{k=1,2,...}$ is a partition of $\mathbb{R}^+$ generated by setting $X_k := f^{-1}(Y_k)$, $\forall k \in \mathbb{Z}$.
- $\Pi^Z := (Z_l)_{l=1,...,L} = \Pi_1 \cup \Pi_2$
- $t_l := \inf\left\{t \in Z^l\right\}$
- $\hat{w}_l := \tilde{w}_{t_l} |Z_l|$
- $G_{\Pi^Y} := \sum_k G_k$, where $G_k := \frac{1}{\phi} \sum_l \hat{w}_l \, y_k \, 1_{\{t_l \in X_k\}}$.

For every $k = 1, ..., L$ we have

$$\begin{aligned}
\mathbb{E}[G_k] &= \mathbb{E}\left[\frac{1}{\phi} \sum_l \hat{w}_l \, y_k \, 1_{\{t_l \in X_k\}}\right] \\
&= y_k \, \mathbb{E}\left[\frac{\sum_l \hat{w}_l \, 1_{\{t_l \in X_k\}}}{\sum_l \hat{w}_l \, 1_{\{t_l \in X\}}}\right] \\
&= y_k \, \frac{|X_k|}{|X|} \\
&= y_k \, |X_k| \quad ,
\end{aligned}$$

where we have used the fact that $\hat{w}_l$ are identically distributed for all $l = 1, ..., L$. Also

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(G_k) &= \mathbb{V}\mathrm{ar}\left[\frac{1}{\phi} \sum_l \hat{w}_l \, y_k \, 1_{\{t_l \in X_k\}} - y_k \, |X_k|\right] \\
&= y_k^2 \, \mathbb{V}\mathrm{ar}\left[\frac{\sum_l \hat{w}_l \, 1_{\{t_l \in X_k\}}}{\sum_l \hat{w}_l \, 1_{\{t_l \in X\}}} - |X_k|\right] \to 0 \;\text{ as }\; M \to 0 \;,
\end{aligned}$$

by the strong law of large numbers. Therefore $G_k \to y_k \, |X_k|$ in L^2.

Given that G_{Π^Y} is equivalent to the simple function used in the definition of the Lebesgue integral, we can show (notice that $\theta = \int_0^1 \rho_t \, \sigma_{1,t} \sigma_{1,t} \mathrm{d}t = \int_0^1 f_t \mathrm{d}t$) that

$$\lim_{\substack{M \to 0 \\ N \to 0}} \mathbb{E}\left[\left(\sum_k G_k - \theta\right)^2\right] = 0$$

The same can be said for

$$\lim_{\substack{M\to 0\\ N\to 0}} \mathbb{E}\left[\left(\sum_k \hat{G}_k - \theta\right)^2\right] = 0 \quad ,$$

where $\hat{G}_i := \frac{1}{\phi}\sum_l \hat{w}_l\, y_{k-1}\, 1_{\{t_l\in X_k\}}$. Given that for every k, as $M\to 0$, we have

$$y_{k-1}\, 1_{\{t\in X_k\}} \leq \left\{\sum_l \left(\frac{1}{|Z_l|}\int_{Z_l} f_t \mathrm{d}t\right) 1_{\{t\in Z_l\}}\right\} 1_{\{t\in X_k\}} \leq y_k\, 1_{\{t\in X_k\}}$$

almost everywhere (in case f_t is discontinuous in x_k (x_{k-1}), the inequality might not hold for $t = x_k$ ($t = x_{k-1}$)), it follows that

$$\lim_{\substack{M\to 0\\ N\to 0}} \mathbb{E}\left[\left(\sum_k H_k - \theta\right)^2\right] = 0 \quad ,$$

where $H_k := \frac{1}{\phi}\sum_l \hat{w}_l \left(\frac{1}{|Z_l|}\int_{Z_l} f_t \mathrm{d}t\right) 1_{\{t\in X_k\}} = \frac{1}{\phi}\sum_l \tilde{w}_{t_l}\left(\int_{Z_l} \rho_t\, \sigma_{1,t}\sigma_{2,t}\mathrm{d}t\right) 1_{\{t_l\in X_k\}}$.

Given that $\sum_k H_k = \frac{1}{\phi}\sum_{i,j} c\left(I_i^1 \cap I_j^2\right) w_{ij}$ (irrespectively of Π^X), we have thus proven that

$$\lim_{M\to 0} \mathbb{E}\left[\left(\frac{1}{\phi}\sum_{i,j} c\left(I_i^1 \cap I_j^2\right) w_{ij} - \theta\right)^2\right] = 0$$

Therefore we have

$$\mathbb{E}\left[D_4\right] = \theta^2 + o(1) \ .$$

Concluding:

$$\mathbb{E}\left[\mathrm{LW}^2\right] = \mathbb{E}\left[D_1 + D_2 + D_3 + D_4\right] = \theta^2 + o(1) \ .$$

2.7.3 Non-Zero Drift Case

Now we consider the case with non-zero drift such that $\sup_{0\leq t\leq T}|\mu_{k,t}| \in L^4$, $k = 1, 2$. Let $N_\cdot^k := \int_0^\cdot \mu_{k,t}\,\mathrm{d}t$, $S_\cdot^k := \int_0^\cdot \sigma_{k,t}\,\mathrm{d}W_{k,t}$, $k = 1, 2$, and

$$B_0 := \frac{T}{\phi}\sum_{i,j}\Delta S_{1,i}\Delta S_{2,j}w_{i,j}\ , \quad B_1 := \frac{T}{\phi}\sum_{i,j}\Delta N_{1,i}\Delta S_{2,j}w_{i,j}$$

$$B_2 := \frac{T}{\phi}\sum_{i,j}\Delta S_{1,i}\Delta N_{2,j}w_{i,j}\ , \quad B_3 := \frac{T}{\phi}\sum_{i,j}\Delta N_{1,i}\Delta N_{2,j}w_{i,j}$$

Note that

$$\begin{aligned}\mathbb{E}\left[B_1\right] &= \mathbb{E}\left[\frac{T}{\phi}\sum_i\int_{I_i^1}\mu_{1,t}\mathrm{d}t\left(\sum_j\int_{I_j^2}\sigma_{2,t}\,\mathrm{d}W_{2,t}\,w_{i,j}\right)\right]\\ &= \mathbb{E}\left[\frac{T}{\phi}\sum_i\int_{I_i^1}\mu_{1,t}\mathrm{d}t\left(\sum_j\mathbb{E}\left\{\int_{I_j^2}\sigma_{2,t}\,\mathrm{d}W_{2,t}\middle|\,\Pi\right\}\,w_{i,j}\right)\right] = 0\ .\end{aligned}$$

Also

$$\begin{aligned}\mathbb{E}\left[B_1^2\right] &= \mathbb{E}\left[\left(\frac{T}{\phi}\sum_{i,j}\Delta N_{1,i}\Delta S_{2,j}w_{i,j}\right)^2\right]\\ &= \mathbb{E}\left[\frac{T^2}{\phi^2}\sum_{i,j,i',j'}\mathbb{E}\left\{\Delta N_{1,i}\Delta S_{2,j}\Delta N_{1,i'}\Delta S_{2,j'}\middle|\,\Pi\right\}w_{ij}w_{i'j'}\right]\\ &\le T^2\mathbb{E}\left[\sup_{0\le t\le T}\left|\mu_{1,t}\right|^2\cdot\frac{1}{\phi^2}\sum_{j,i,i'}\left|I_i^1\right|\left|I_{i'}^1\right|\mathbb{E}\left\{\Delta S_{2,j}^2\middle|\,\Pi\right\}w_{ij}w_{i'j}\right]\\ &= T^2\mathbb{E}\left[\sup_{0\le t\le T}\left|\mu_{1,t}\right|^2\cdot\frac{1}{\phi^2}\sum_{j,i,i'}\left|I_i^1\right|\left|I_{i'}^1\right|c_2(I_j^2)w_{ij}w_{i'j}\right]\\ &\le T^2\sup_{0\le t\le T}\left(\sigma_{2,t}\right)^2\cdot\mathbb{E}\left[\sup_{0\le t\le T}\left|\mu_{1,t}\right|^2\cdot\frac{1}{\phi^2}\sum_{j,i,i'}\left|I_i^1\right|\left|I_{i'}^1\right|\left|I_j^2\right|w_{ij}w_{i'j}\right]\\ &\le \sup_{0\le t\le T}\left(\sigma_{2,t}\right)^2\cdot\mathbb{E}\left[\sup_{0\le t\le T}\left|\mu_{1,t}\right|^2\cdot 16M^2\sum_{j,i,i'}\left|I_i^1\cap I_j^2\right|\frac{\left|I_{i'}^1\cap I_j^2\right|}{\left|I_j^2\right|}\right]\\ &= \sup_{0\le t\le T}\left(\sigma_{2,t}\right)^2\cdot\mathbb{E}\left[\sup_{0\le t\le T}\left|\mu_{1,t}\right|^2\cdot 16M^2\sum_{i,j}\left|I_i^1\cap I_j^2\right|\right]\\ &= \sup_{0\le t\le T}\left(\sigma_{2,t}\right)^2\cdot\mathbb{E}\left[\sup_{0\le t\le T}\left|\mu_{1,t}\right|^2\cdot 16M^2T\right]\ ,\end{aligned}$$

where in the last inequality we have applied (2.7.2). Given that the supremum of μ_1 is in L^4, it follows that $\mathbb{E}[B_1^2] = o(1)$. $\mathbb{E}[B_2^2] = o(1)$ and $\mathbb{E}[B_3^2] = o(1)$ can be shown similarly.

Because it holds that [1]

$$\mathbb{E}\left[(\mathrm{LW}_n - \theta)^2\right] \leq 2\,\mathbb{E}\left[(B_0 - \theta)^2\right] + 6\,\mathbb{E}\left[B_1^2 + B_2^2 + B_3^2\right] \ ,$$

it follows that, as $M \to 0$, $\mathbb{E}\left[(\mathrm{LW}_n - \theta)^2\right] = 0$.

2.8 Simulation Results

Table 2.1 and Figure 2.1 summarize the efficiency improvements of the Linearly Weighted covariance estimator compared to the one of Hayashi and Yoshida [2005]. In particular, to produce the results displayed in Table 2.1 we simulated 100,000 sample paths, over the interval $[0, 1]$, of two diffusion processes, $X^{(1)}$ and $X^{(2)}$, with initial value $X_0^{(1)} = 0$ and $X_0^{(2)} = 0$, drift $\mu_{(1)} = \mu_{(2)} = 0$ and volatility parameters $\sigma^2_{(1)} = 1$ and $\sigma^2_{(2)} = 1$. At the start of each simulation we drew the correlation parameter $\rho_{1,2}$ from a uniform distribution $U(-1, 1)$ and generated the observation times of the two processes by applying a Poisson random sampling scheme, whereby the set of observation times is represented by the jump times of two independent Poisson processes with intensity $\lambda_{(1)} = 55$ and $\lambda_{(2)} = 55$ respectively over the same simulation interval $[0, 1]$.

Table 2.1: LW vs. HY

	Average Error	RMSE	Relative Efficiency
H.Y.	-0.0018	0.3025	125.40%
L.W.	-0.0017	0.2413	100.00%

[1] The following result has been used: let $a, b, c, d \in \mathbb{R}$, then

$$\begin{aligned}
[(a+b+c+d)-\theta]^2 &= (a-\theta)^2 + (b+c+d)^2 + 2\,(a-\theta)\,(b+c+d) \\
&\leq 2\,(a-\theta)^2 + 2\,(b+c+d)^2 \\
&= 2\,(a-\theta)^2 + 2\left(b^2+c^2+d^2+2bc+2bd+2cd\right) \\
&\leq 2\,(a-\theta)^2 + 6\left(b^2+c^2+d^2\right)
\end{aligned}$$

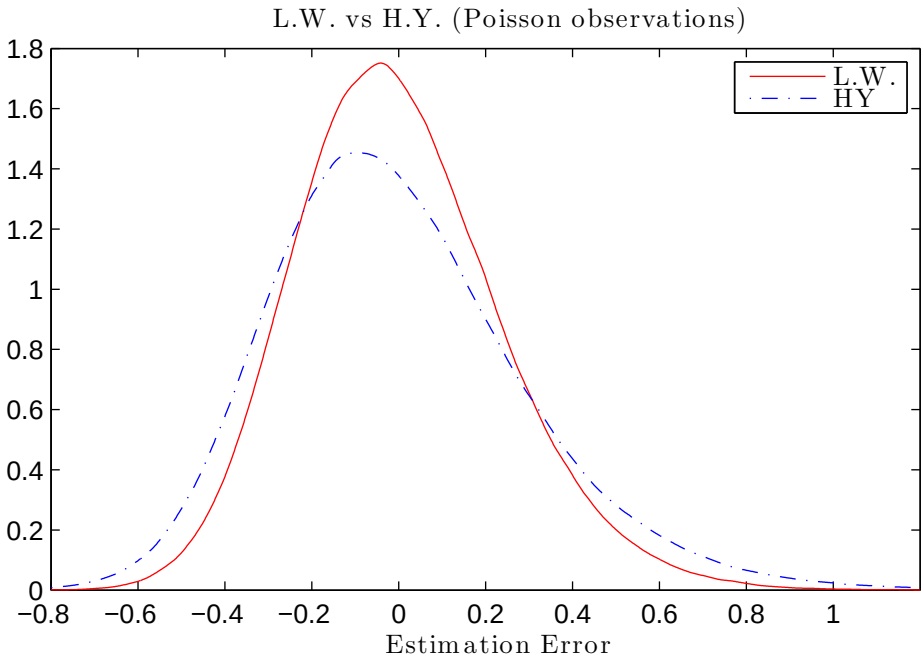


Figure 2.1: Empirical densities (by kernel density estimation) of LW, HY using Poisson observations. Simulation parameters: $\sigma^2_{(1)} = 1$, $\sigma^2_{(2)} = 1$ and $\rho_{1,2} = 0.5$

Figure 2.1 displays the empirical densities by using the same simulation approach as above, except for the fact that the correlation parameter $\rho_{1,2}$ has been kept constant at each simulation with a value of 0.5.

2.9 Conclusion

In this chapter we have introduced a new unbiased covariance estimator for diffusion processes that are observed over a non-synchronous time grid - we call it the "Linearly Weighted" estimator (LW) - which takes advantage of all available observations (tick-by-tick) without requiring any pre-processing or subsampling of the dataset. The LW estimator, shares the simplicity of implementation and low computational cost of Hayashi and Yoshida [2005] but, on the other hand, it achieves significantly higher efficiency (around 20 % reduction in RMSE in case of Poisson sampling) by taking into account the level of overlapping between each observation, thereby applying a higher weight when the intervals closely overlap,

resulting in a less 'noisy' sample, and lower weight when the overlap is limited.

We also analysed LW in the QMLE context of the numerical estimation approach (KEM) recently introduced in Corsi et al. [2012] and Shephard and Xiu [2012], noting that LW can be seen as the result of a similar iterative EM procedure based on a sub-optimal predictor for the missing values on the synchronous grid represented by the union of all available observation times, and we showed how this difference leads to an estimator which is much simpler to implement and fast to compute (since it is available in closed form), while, at the same time, preserving most of the efficiency gains (cf. Chapter 3) of KEM.

Finally, we proved the unbiasedness and consistency of the LW estimator under time-dependent (deterministic) covariance and assuming stationary distributed observation times.

Chapter 3

Multivariate Realised Covariance from Nonsyncronous Timeseries Using the Regularized Linearly Weighted Estimator

The Linearly Weighted (LW) covariance estimator introduced in Chapter 2 provides a simple, unbiased and computationally efficient way of estimating the covariance between two diffusion processes observed over a non synchronous grid. In this chapter we show how the LW estimator is particularly suitable to be applied together with the regularization methodology suggested by Rebonato and Jackel [2000] to produce positive semi-definite estimates of multivariate realized covariance matrices that do not require 'synchronization' processing of the original data. Simulation studies show that the regularization procedure (when required) has a positive impact on the quality of the estimate and compare the performance of the proposed estimator with alternatives from the literature. In addition, we show how the LW estimator can be enhanced to make it robust to noisy data and introduce an unbiased version of the estimator to be used whenever the drifts of the processes are estimated in sample as well as a generalization of the LW estimator to Ornstein-Uhlenbeck processes.

3.1 Introduction

Estimating the multivariate covariance of financial variables based on time series data is a common requirement for a wide range of financial applications, including portfolio selection, risk management and forecasting. In all of these contexts, practitioners have to deal with the challenge that most real-world datasets don't satisfy the assumptions of synchronicity of the observations, even-spacing and completeness of the most widely applied estimators (such as the 'Realized Covariance' estimator described in Chapter 2). In the case of risk management applications, where analysis are often based on daily frequency data, dataset issues may arise because of time series having heterogeneous sampling frequencies (e.g. while equity and interest rate time series are available at daily frequency, time series for real estate, inflation or hedge fund cumulated returns are typically released on a monthly schedule) and/or different publication days (e.g. one monthly time series has new data points every month's end while another one has mid-month schedule), because of missing observations produced by a multitude of national calendars or simply because of corruption of data. Synchronicity assumptions become even more unrealistic when estimates are obtained using high frequency data. One common approach in these cases has been to apply either previous-tick or linear interpolation to the missing data, however, as it was shown in Hayashi and Yoshida [2005] and, more recently, in Zhang [2011], these methods based on the synchronization of non synchronous data lead to biased estimates. Various other approaches have been proposed in the literature to tackle this asynchronicity problem: incorporate lead and lag cross returns in the estimator (Scholes and Williams [1977]; Cohen et al. [1983]; Bollerslev and Zhang [2003]; Bandi and Russell [2005]), avoid any synchronization by directly using tick-by-tick data (De Jong and Nijman [1997]; Hayashi and Yoshida [2005]; Palandri [2006]; Sheppard [2006]), adopt the so called refresh time scheme (Barndorff-Nielsen et al. [2011]; Aït-Sahalia et al. [2010]; Zhang [2011]), and the multivariate Fourier method (Renò [2003]; Mancino and Sanfelici [2011]).

In this chapter we consider applications at both high and low frequency, the difference being that at high frequencies time series can generally be assumed as having no drift (being of order dt, the drift is mathematically negligible compared

to the diffusive component which is of order $\sqrt{\mathrm{d}t}$) while at lower frequencies it may be desirable to estimate and then remove the drift from the sample before applying the estimator, with the consequence of introducing a bias to the estimator which need to be evaluated and corrected. In section 3.4.2 we highlight a very useful property (verified empirically) of the Linearly Weighted (LW) estimator that we introduced in Chapter 2, whereby the variance of the estimator is very closely linked to the value of its denominator ϕ. By taking advantage of this property, we extend the application of LW to the estimation of covariance matrices. In particular, we show how the LW estimator, which is not guaranteed, on its own, to produce positive semi-definite estimates, is especially suitable to be used together with the methodology presented in Rebonato and Jackel [2000] which allows to perform the regularization of a non positive semi-definite correlation estimate by carrying out, given a measure of the 'distance' from the original matrix, an unconstrained optimization expressed in terms of angle vectors describing coordinates on a unit hypersphere. Simulation results show that the regularization procedure performed using the recommended weights, based on the value of ϕ, has (when it is necessary) a positive impact on the RMSE of the estimate. In order to be able to apply the suggested approach, in section 3.4.1 we introduce a new correlation estimator based on LW which, in turn, allows us to define a new 2-step variant of the LW estimator of covariance displaying further efficiency improvements compared to the one introduced in Chapter 2.

In the rest of the chapter we present several other results related to the new estimator. Specifically, in section 3.5 we introduce the bias-corrected version of the LW estimator for the case when the drift of the process is assumed to be constant and estimated in sample and, in section 3.6, we generalize the LW estimator to Ornstein-Uhlenbeck processes. Finally, in section 3.7, we consider the potential impact of microstructure noise on the estimator and we derive, under the assumption of knowledge of the univariate parameters describing the variance of noise and that of the processes under investigation, an estimation method which aims at finding a sub-sampling of the dataset that minimizes the RMSE of the estimate through the maximization of the denominator ϕ of the estimator. In conclusion, simulation results in section 3.8 give examples of the performance of LW in several contexts showing how the new estimator has the

benefit of performing closely to the maximum likelihood estimator (which is only available through computationally intensive - particularly in high dimensions - numerical procedures) while at the same time being very simple to implement and fast to compute even for high-dimensional datasets.

3.2 Regularization Through Hypersphere Decomposition

As summarized in Rapisarda et al. [2007], correlation matrices must satisfy four basic properties in order to bear statistical and financial significance:

1. all their entries must lie in the interval $[-1, 1]$;
2. diagonal entries must be equal to one;
3. the matrix must be symmetric (correlation between variables a and b is equal to correlation between b and a);
4. the matrix must be 'positive semi-definite'. This has to do with the fact that the variance of a portfolio P, whose correlation matrix is C, is $\sigma_P^2 = \mathbf{w}^\mathsf{T} C \mathbf{w} \geq 0$, where $\mathbf{w}$ is the array of weights of the constituent financial variables, each multiplied by the standard deviation of the respective variable.

Rebonato and Jackel [2000] introduced a regularization method for a generic $n \times n$ correlation matrix C and user-defined error measure ε based on the standard angles parameterization. Specifically, let $\mathbf{C}$ denote the set of all $n \times n$ correlation matrices satisfying properties 1-4 and let $\hat{C} \in \mathbf{C}$ be such that $\varepsilon := \|C - \hat{C}\|$ is minimized. Applying the well-known result from linear algebra that every symmetric positive semi-definite matrix M can be decomposed as the product

$$M = WW^\mathsf{T},$$

they decompose the, yet unknown, regular covariance matrix $\hat{C}$ as

$$\hat{C} = BB^\mathsf{T}. \tag{3.2.1}$$

Then, they show that the elements of each row vector of matrix B can be seen as the Cartesian coordinates of a point lying on a n-dimensional unit hyper-sphere. Based on this, they propose to parametrize the elements b_{ij} of the matrix B in terms of $n \times (n-1)$ angular coordinates θ_{ij}. Specifically:

$$b_{ij} = \begin{cases} \cos\theta_{ij} \prod_{k=1}^{j-1} \sin\theta_{ik}, & \text{for } j < n, \\ \prod_{k=1}^{n-1} \sin\theta_{ik}, & \text{for } j = n, \end{cases} \tag{3.2.2}$$

for $i = 1, ..., n$, where we have $b_{i1} = \cos\theta_{i1}$. This parametrization facilitates the identification of $\hat{C}$ by allowing the use of unconstrained optimization.

As we will see in the following sections, the LW estimator provides, as a by-product, a good proxy for the variance of its correlation estimates c_{ij}. The availability of this information will enable us to apply the regularization procedure based on standard angles parametrization (3.2.2) using a 'weighted sum of squares' error measure. Specifically

$$\varepsilon_W := \sum_{l,k} W_{ij} \left(c_{ij} - \hat{c}_{ij}\right)^2 \ , \tag{3.2.3}$$

where c_{ij} and $\hat{c}_{ij}$ are elements of matrix C and $\hat{C}$ and where

$$W_{ij} \approx \frac{1}{\mathbb{V}\mathrm{ar}[c_{ij}]} \ .$$

In other words, by minimizing ε_W we can ensure that the elements of C that have been estimated with highest accuracy will be left virtually untouched by regularization while the less reliable values (which are most likely to determine a non-PSD C) are more freely modified.

3.3 Problem setup

In this chapter we consider a d-dimensional diffusion process $X = \left(X^{(1)}, \ldots, X^{(d)}\right)'$ defined by the following stochastic differential equation

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \mathrm{S}_t \mathrm{d}W_t \ , \tag{3.3.1}$$

where W is a d-dimensional vector of correlated Brownian motions such that $\mathrm{d}\langle W_t, W_t\rangle = \mathrm{C}_t\,\mathrm{d}t$, μ_t is a vector of elements which are predictable locally bounded drifts and the diffusion coefficient S_t is a $d\times d$ diagonal matrix whose elements are deterministic and bounded functions. The values of $\left(X^{(1)},\ldots,X^{(d)}\right)'$ are observed in the interval $[0,T]$ at times that are irregularly spaced and non synchronous. We use $\Pi_l := \left(t_j^{(l)}\right)_{j=1,\ldots,N^{(l)}}$ to denote the available observation times for asset l, $l=1,\ldots d$ and we refer to the time intervals between successive observations as $I_j^{(l)} := \left(t_{j-1}^{(l)}, t_j^{(l)}\right]$, with $j=1,\ldots,N^{(l)}$. Furthermore, we write the union of all observation times as

$$\Pi = (t_i)_{i=1,\ldots,n} = \bigcup_{l=1}^{d} \Pi_l, \qquad i=1,2,\ldots,n.\ ,$$

where times have been ordered so that $0 \le t_1 < \cdots < t_i < \cdots < t_n \le T$.

3.4 LW Estimator

The Linearly Weighted covariance estimator introduced in Chapter 2 can be extended to the multivariate context by applying the estimator to each pair of assets. Specifically:

Definition 3.4.1 (LW estimator)**.**

$$\mathrm{LW}_{l,k} \equiv \mathrm{LW}\left(X^{(l)}_{\Pi_l}, X^{(k)}_{\Pi_k}\right) := \frac{1}{\phi^{(l,k)}} \sum_{i,j} \Delta X_i^{(l)} \Delta X_j^{(k)} w_{i,j}^{(l,k)}\ , \tag{3.4.1}$$

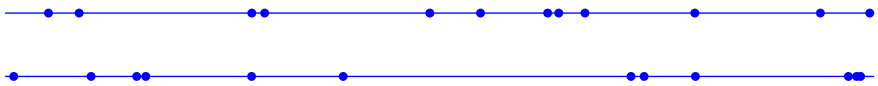
with

$$\phi^{(l,k)} := \sum_{i,j} \left| I_i^{(l)} \cap I_j^{(k)} \right| w_{i,j}^{(l,k)}\ ,$$

$$w_{i,j}^{(l,k)} := \frac{\left| I_i^{(l)} \cap I_j^{(k)} \right|}{\left| I_i^{(l)} \right| \left| I_j^{(k)} \right|}\ ,$$

where $I_i^{(l)} := \left(t_{i-1}^{(l)}, t_i^{(l)}\right]$, $i=1,\ldots,N(l)$, $l=1,\ldots,d$, and where the measure $|\cdot|$ is used to denote the length of an interval (i.e. $|I_i^{(l)}| := t_i^{(l)} - t_{i-1}^{(l)}$).

Original data:



Processed data:

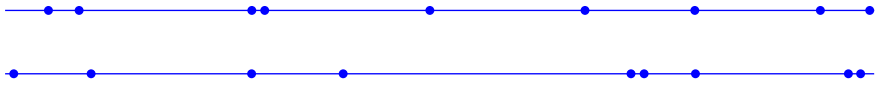


Figure 3.1: Processing of the observations

3.4.1 Correlation Estimator

The LW covariance estimator can be used to obtain a consistent estimate of the correlation between two assets in a straightforward way by 'standardizing' $\mathrm{LW}_{l,k}$ as follows.

Definition 3.4.2 (LW correlation estimator)**.** Let $l,k=1,\ldots,d$, $i=1,\ldots,N(l)$ and $j=1,\ldots,N(k)$. Then:

$$R_{l,k} := \frac{\frac{1}{\phi^{(l,k)}}\sum_{i,j}\Delta X_i^{(l)}\Delta X_j^{(k)}w_{i,j}^{(l,k)}}{\sqrt{\frac{1}{\phi^{(l,l)}}\sum_l\left(\Delta X_i^{(l)}\right)^2 w_{i,i}^{(l,l)}}\sqrt{\frac{1}{\phi^{(k,k)}}\sum_k\left(\Delta X_j^{(k)}\right)^2 w_{j,j}^{(k,k)}}} = \frac{\mathrm{LW}_{l,k}}{\sqrt{\mathrm{LW}_{l,l}}\sqrt{\mathrm{LW}_{k,k}}}\,. \tag{3.4.2}$$

However, the correlation estimator can be made more efficient by introducing changes to the way we calculate the volatility estimates $\sqrt{\mathrm{LW}_{l,l}}$ and $\sqrt{\mathrm{LW}_{k,k}}$. Let us define the following data pre-processor:

Pre-processor (A)**.** Discard all observations $t_i^{(l)}$ for which there exists j such that $\left(I_i^{(l)}\cup I_{i+1}^{(l)}\right)\subseteq I_j^{(k)}$ and observations $t_j^{(k)}$ for which there exists i such that $\left(I_j^{(k)}\cup I_{j+1}^{(k)}\right)\subseteq I_i^{(l)}$ (see Figure 3.1).

The modified correlation estimator $\widetilde{R}_{l,k}$ is obtained by first applying Pre-processor A to the dataset and then calculating the following estimator.

Definition 3.4.3 (LW correlation estimator - modified)**.** Let $\widetilde{\Pi}_l^{(l,k)}$ and $\widetilde{\Pi}_k^{(l,k)}$ be a set of observation times resulting from applying Pre-processor A to Π_l and Π_k. Then we define

$$\begin{aligned}\widetilde{R}_{l,k} \equiv \widetilde{R}_{l,k}\left(X^{(l)}_{\widetilde{\Pi}_l^{(l,k)}}, X^{(k)}_{\widetilde{\Pi}_k^{(l,k)}}\right) :=& \frac{\frac{1}{\phi^{(l,k)}}\sum_{i,j}\Delta X_i^{(l)}\Delta X_j^{(k)} w_{i,j}^{(l,k)}}{\sqrt{\frac{1}{\widetilde{\phi}^{(l)}}\sum_{i,j}\left(\Delta X_i^{(l)}\right)^2 \widetilde{w}_{i,j}^{(l,k),i}}\sqrt{\frac{1}{\widetilde{\phi}^{(k)}}\sum_{i,j}\left(\Delta X_j^{(k)}\right)^2 \widetilde{w}_{i,j}^{(l,k),j}}} \\ =& \frac{\mathrm{LW}_{l,k}}{\sqrt{\widetilde{\mathrm{LW}}_{l,l}}\sqrt{\widetilde{\mathrm{LW}}_{k,k}}} ,\end{aligned} \tag{3.4.3}$$

where

$$\widetilde{w}_{i,j}^{(l,k),i} := w_{i,j}^{(l,k)} \cdot \frac{\left|I_i^{(l)} \cap I_j^{(k)}\right|}{\left|I_i^{(l)}\right|} ,$$

$$\widetilde{w}_{i,j}^{(l,k),j} := w_{i,j}^{(l,k)} \cdot \frac{\left|I_i^{(l)} \cap I_j^{(k)}\right|}{\left|I_j^{(k)}\right|} ,$$

and

$$\widetilde{\phi}^{(l)} := \sum_{i,j}\left|I_i^{(l)}\right| \widetilde{w}_{i,j}^{(l,k),i} ,$$

$$\widetilde{\phi}^{(k)} := \sum_{i,j}\left|I_j^{(k)}\right| \widetilde{w}_{i,j}^{(l,k),j} .$$

Note that the removal of observations based on Pre-processor A affects only the values of the variance estimators $\widetilde{\mathrm{LW}}_{l,l}$ and $\widetilde{\mathrm{LW}}_{k,k}$ while leaving $\mathrm{LW}_{l,k}$ unaffected. The intuition behind this change in the variance estimators is that we now use weights that are consistent with the ones used by the LW estimator of covariance. Furthermore, given that, by construction, $\phi^{(l,k)} = \widetilde{\phi}^{(l)} = \widetilde{\phi}^{(k)}$, the computation of $\widetilde{R}_{l,k}$ simplifies to

$$\widetilde{R}_{l,k} := \frac{\sum_{l,k}\Delta X_i^{(l)}\Delta X_j^{(k)} w_{l,k}^{(l,k)}}{\sqrt{\sum_{l,k}\left(\Delta X_i^{(l)}\right)^2 \widetilde{w}_{l,k}^{(l,k),i}}\sqrt{\sum_{l,k}\left(\Delta X_j^{(k)}\right)^2 \widetilde{w}_{l,k}^{(l,k),j}}} . \tag{3.4.4}$$

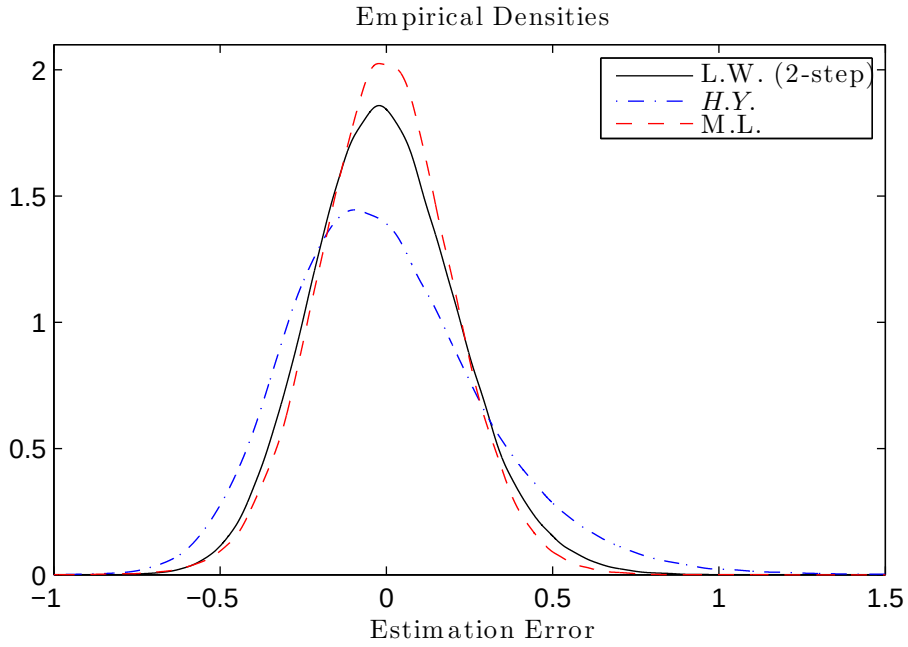


Figure 3.2: Empirical densities (by kernel density estimation) of LW, HY and ML covariance estimators, using Poisson observations and 100,000 samples. Simulation parameters: $\sigma^2_{(1)} = 1$, $\sigma^2_{(2)} = 1$ and $\rho_{1,2} = 0.5$.

Irrespective of which version of the estimator is used, $\widetilde{R}_{l,k}$ is not guaranteed to be bounded by 1 in magnitude. In case no regularization procedure is applied to the estimated correlation matrix $\widetilde{C}$, following Hayashi and Yoshida [2005] we can apply to $\widetilde{R}_{l,k}$ a projection transformation from $\mathbb{R}$ to $[-1, 1]$. Specifically

$$\hat{R}_{l,k} = \Pi_{[-1,1]}\left(\widetilde{R}_{l,k}\right)$$

where $\Pi_{[-1,1]}(x)$ is the closest point in $[-1, 1]$ to x. If, on the other hand, the estimated correlation matrix $\widetilde{C}$ is being regularized using a procedure such as the one described in section 3.2, then this step is not necessary since all elements of the regularized matrix are, by construction, bounded between -1 and 1.

Concluding this section we note that it is possible to take advantage of the efficiency improvements brought about by the correlation estimator $\widetilde{R}_{l,k}$ to further increase the efficiency of the LW estimator of covariance. Specifically, by using the following:

Definition 3.4.4 (LW 2-step estimator)**.**

$$\mathrm{LW}_{l,k}^{2-\mathrm{step}} := \widetilde{R}_{l,k} \sqrt{\mathrm{LW}_{l,l}} \sqrt{\mathrm{LW}_{k,k}} \ ,$$

Remark 3.4.5. Note that $\mathrm{LW}_{l,l}$ coincides with the Maximum Likelihood estimator of the variance of $X^{(k)}$ under the assumption of constant volatility (QML estimator in case volatility is not constant).

The empirical densities in Figure 3.2 show that the LW 2-step estimator significantly outperforms the Hayashi-Yoshida estimator and, in fact, it has a distribution that is very close to the one of the maximum likelihood estimator (which was calculated numerically).

3.4.2 ϕ and variance of the estimator

Empirical results show that the variance of the LW estimators of covariance (3.4.1) and correlation (3.4.4) are very close to proportional to $\frac{1}{\phi^{(l,k)}}$, i.e. $\mathbb{V}\mathrm{ar}\left[\mathrm{LW}_{l,k}\right] \propto \frac{1}{\phi^{(l,k)}}$ and $\mathbb{V}\mathrm{ar}\left[\widetilde{R}_{l,k}\right] \propto \frac{1}{\phi^{(l,k)}}$. In particular, we can see in Figure 3.3 and 3.4 and in Table 3.1 that, as we draw different sets of random observation times, the value of ϕ always gives accurate information about the relative variance of the estimates provided by the LW estimator. This property makes $\phi^{(l,k)}$ a natural choice

Table 3.1: Explained variance

	R^2 of $\phi^{(l,k)}$
$\mathbb{V}\mathrm{ar}\left[\mathrm{LW}_{l,k}\right]$	0.993739066
$\mathbb{V}\mathrm{ar}\left[\widetilde{R}_{l,k}\right]$	0.985423818

as weight in the objective function (3.2.3). In other words, we can apply the regularization procedure of Rebonato and Jackel [2000] knowing which entries of the correlation matrix are estimated with more and less accuracy, thereby leaving the most reliable estimates virtually unchanged while allowing more pronounced adjustments where values are estimated with larger statistical errors. Specifically, we can regularize matrix C based on the following weighted sum of squares error

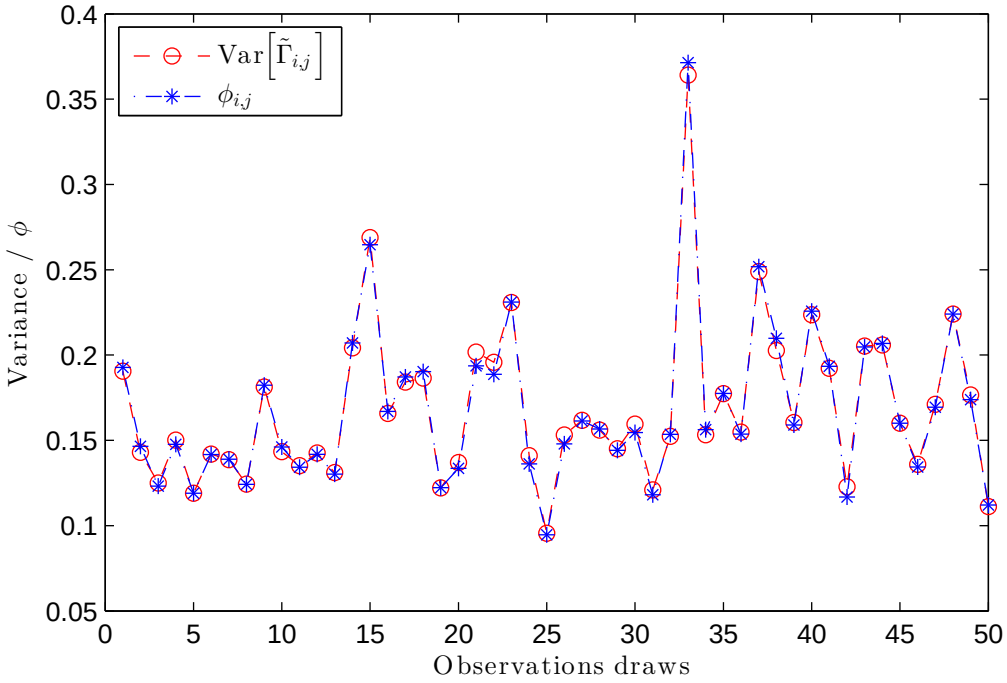


Figure 3.3: Proportionality between $\mathbb{V}\mathrm{ar}\left[\mathrm{LW}_{l,k}\right]$ and $\frac{1}{\phi^{(l,k)}}$ through 50 different sets of Poisson observations.

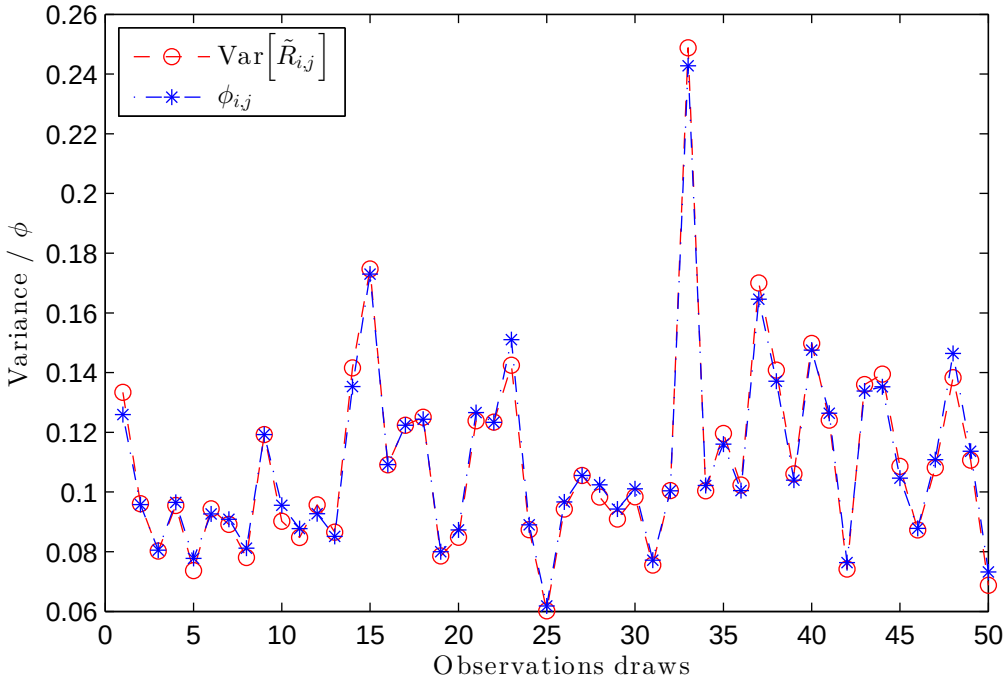


Figure 3.4: Proportionality between $\mathbb{V}\mathrm{ar}\left[\widetilde{R}l,k\right]$ and $\frac{1}{\phi^{(l,k)}}$ through 50 different sets of Poisson observations.

metric:

$$\varepsilon_\phi := \sum_{l,k} \phi^{(l,k)} \left(c_{ij} - \hat{c}_{ij}\right)^2 \ . \tag{3.4.5}$$

3.5 Estimation with Drift

When working outside of the high-frequency context the drift of the data generating process cannot be assumed to be negligible and it may therefore be appropriate to estimate it so it can be subtracted from the observed increments. Doing so, however, has the effect of introducing bias to the estimate due to the reduction in the degrees of freedom of the estimator. In this section we introduce the bias-corrected version of the LW estimator for the case when the drift of the process is assumed to be deterministic and estimated in sample. Continuing with the QMLE approach introduced in Chapter 2 (and also found in Xiu [2010] and Shephard and Xiu [2012]), we derive the new bias-corrected estimator starting from a mis-specified version of the model where drift and diffusion coefficients are assumed to be constant.

Definition 3.5.1 (LW Estimator - Non Zero Mean)**.** Let $|I|$ to denote the length of interval I. Then the bias-corrected LW estimator to be applied when the drift of the process is estimated in sample is given by

$$\text{LW}_{\bar{\mu}} := \frac{1}{\phi - \eta} \sum_{i,j} \left(\Delta X_i^{(1)} - \left|I_i^{(1)}\right| \ \bar{\mu}_X\right) \left(\Delta X_j^{(2)} - \left|I_j^{(2)}\right| \ \bar{\mu}_Y\right) w_{ij} \ ,$$

where

$$w_{ij} := \frac{\left|I_i^{(1)} \cap I_j^{(2)}\right|}{\left|I_i^{(1)}\right| \left|I_j^{(2)}\right|}, \quad \phi := \sum_{i,j} w_{ij} \left|I_i^{(1)} \cap I_j^{(2)}\right| \ ;$$

and the bias-correction term η is defined as

$$\begin{aligned} \eta :=& \frac{1}{|I|} \left(|I \cap J| - \beta^S \left(1 - \beta^S\right) \ |J_1| - \beta^E \left(1 - \beta^E\right) \ |J_M|\right) \\ &+ \frac{1}{|J|} \left(|I \cap J| - \alpha^S \left(1 - \alpha^S\right) \ |I_1| - \alpha^E \left(1 - \alpha^E\right) \ |I_N|\right) \\ &- \frac{|I \cap J|^2}{|I| \ |J|} \ , \end{aligned}$$

with

$$\alpha^S := \big(t_0^X \vee t_0^Y - t_0^X\big)/|I_1|, \quad \alpha^E := \big(t_N^X \wedge t_M^Y - t_{N-1}^X\big)/|I_N|, \ ,$$
$$\beta^S := \big(t_0^X \vee t_0^Y - t_0^Y\big)/|J_1|, \quad \beta^E := \big(t_N^X \wedge t_M^Y - t_{M-1}^Y\big)/|J_M| \ ;$$

while the drift estimators $\bar{\mu}_X$ and $\bar{\mu}_Y$ are as follows

$$\bar{\mu}_X := \frac{1}{|I|}\sum_i \Delta X_i^{(1)}, \quad \bar{\mu}_Y := \frac{1}{|J|}\sum_j \Delta X_j^{(2)} \ .$$

Lemma 3.5.2. *If $\mu_t^l = c$, $c \in \mathbb{R}$, $0 \le t \le T$, then $\mathrm{LW}_{\bar{\mu}}$ is unbiased.*

Proof. See Appendix B.1. □

Remark 3.5.3. Note that when the observations of $X^{(1)}$ and $X^{(2)}$ are synchronous and equally spaced, then we have that ϕ is equal to the number of available increments and the bias correction term η is equal to 1, i.e. we obtain the well known sample covariance estimator.

3.6 Generalization to Ornstein-Uhlenbeck Processes

Ornstein-Uhlenbeck models are continuous-time processes which have broad applications in finance such as in the modelling of interest rates, foreign exchange rates and volatility processes (see for example Barndorff-Nielsen and Shephard [2001]). The multivariate Ornstein-Uhlenbeck process we consider here is defined by the following stochastic differential equation

$$\mathrm{d}X_t = \mu\,\mathrm{d}t - \mathrm{K}\,(X_t - a)\,\mathrm{d}t + \mathrm{S}\,\mathrm{d}W_t \ , \tag{3.6.1}$$

where

$$\mathrm{S} = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_d \end{pmatrix} = \mathrm{diag}\,(s) \ , \tag{3.6.2}$$

and where the instantaneous correlation C and covariance matrix Σ are such that

$$\mathrm{C}\,\mathrm{d}t = \mathrm{d}\langle W_t, W_t\rangle \ , \quad \text{and} \quad \Sigma = S\,\mathrm{C}\,S^{\mathsf{T}} \ .$$

The solution to (3.6.1) is given by

$$X_{t+\tau} = X_t\, e^{-\mathrm{K}\,\tau} + \left(\mathbf{I} - e^{-\mathrm{K}\,\tau}\right)\frac{\mu}{\kappa} + \left(\mathbf{I} - e^{-\mathrm{K}\,\tau}\right) a + \int_t^{t+\tau} e^{-\mathrm{K}\,(t+\tau-s)}\,\mathrm{S}\,\mathrm{d}W_s \ .$$

Here we consider only multivariate Ornstein-Uhlenbeck processes where auto-regression/mean-reversion is allowed only component-wise, i.e. each dimension with respect to itself. More rigorously we have

$$\mathrm{K} = \begin{pmatrix} \kappa_1 & 0 & \cdots & 0 \\ 0 & \kappa_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \kappa_d \end{pmatrix} = \mathrm{diag}\,(\kappa) \ .$$

Before attempting to estimate Σ, we transform the process X in order to extract the drift-less process from the available observations. To this end we define

$$Y_t = \left(Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(d)}\right)' \ ,$$

where

$$\begin{aligned} Y^{(l)}_{t_i^{(l)}} :=& \sum_{j=1}^{i} \left\{ X_{t_j^{(l)}} - \mathbb{E}\left[X_{t_j^{(l)}} \middle| X_{t_{j-1}^{(l)}} \right] \right\} \\ =& \sum_{j=1}^{i} \left\{ X_{t_j^{(l)}} - X_{t_{j-1}^{(l)}}\, e^{-\kappa_l \left(t_j^{(l)} - t_{j-1}^{(l)}\right)} - \left(\mathbf{I} - e^{-\kappa_l \left(t_j^{(l)} - t_{j-1}^{(l)}\right)}\right) a \right. \\ & \left. - \left(\mathbf{I} - e^{-\kappa_l \left(t_j^{(l)} - t_{j-1}^{(l)}\right)}\right) \frac{\mu_l}{\kappa_l} \right\} \\ =& s_l \int_{t_0^{(l)}}^{t_i^{(l)}} e^{-\kappa_l\,(u_l(v)-v)}\,\mathrm{d}W_v^{(l)} \ , \qquad i = 1, \ldots, N(l) \ , \end{aligned}$$

with

$$u_l(t) := \inf\left\{t_j^{(l)} : \quad t_j^{(l)} \geq t \ , \ j = 0, \ldots, N(l)\right\} \ ,$$

and, more generally

$$Y_t^{(l)} := s_l \int_{t_0^{(l)}}^{t} e^{-\kappa_l \, (u_l(v)-v)} \, \mathrm{d}W_v^{(l)} \ , \qquad t \leq t_{N(l)}^{(l)} \ . \tag{3.6.3}$$

Therefore the stochastic differential equation of Y is given by

$$\mathrm{d}Y_t^{(l)} = e^{-\kappa_l \, (u_l(t)-t)} s_l \, \mathrm{d}W_t^{(l)} \ , \qquad l = 1, \ldots, N(l) \ .$$

Then the increment of Y between two consecutive observation times is given by

$$\Delta Y_i := Y_{t_i} - Y_{t_{i-1}} = \int_{t_{i-1}}^{t_i} e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(v)-v)} \, \mathrm{S} \, \mathrm{d}W_v \ ,$$

where the $\mathrm{diag}(\cdot)$ operator transforms a d-dimensional vector into a $d \times d$ diagonal matrix having the elements of the vector on the main diagonal, and where

$$\mathbf{u}\,(t) := \left(u_{(1)}(t) \ , \ u_{(2)}(t) \ , \ \ldots \ , \ u_{(d)}(t)\right)' \ .$$

From the definition of Y it follows that

$$\mathbb{E}\left[\Delta Y_i \, \Delta Y_i'\right] = \int_{t_{i-1}}^{t_i} e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(t_i)-t)} \Sigma \left(e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(t_i)-t)}\right)' \mathrm{d}t \ . \tag{3.6.4}$$

Notice that, given that K is a diagonal matrix, we can re-arrange (3.6.4) as

$$\mathbb{E}\left[\Delta Y_i \, \Delta Y_i'\right] = \Sigma \circ \int_{t_{i-1}}^{t_i} e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(t_i)-t)} \mathbf{1}\mathbf{1}' \left(e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(t_i)-t)}\right)' \mathrm{d}t \quad , \tag{3.6.5}$$

where $\circ$ denotes the Hadamard product, i.e. element-wise matrix multiplication, and $\mathbf{1}$ is a d-dimensional vector of ones, from which it follows that

$$\mathbb{E}\left[\left(\Delta Y_i \, \Delta Y_i'\right) \oslash \int_{t_{i-1}}^{t_i} e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(t_i)-t)} \mathbf{1}\mathbf{1}' \left(e^{-\mathrm{K} \, \mathrm{diag}(\mathbf{u}(t_i)-t)}\right)' \mathrm{d}t\right] = \Sigma \quad ,$$

where the notation '$\oslash$' represents element-by-element division. Based on this we can write the log-likelihood function of X under the assumption of complete observations at each point of the ordered union of all observation times $\Pi = (t_i)_{i=1,\dots,n}$. That is

$$\log f\left(X_{1:n};\Sigma\right) := c - \frac{1}{2}\sum_{i=2}^{n}\log\left|\int_{t_{i-1}}^{t_i} e^{-\mathrm{K}\,\mathrm{diag}(\mathbf{u}(t_i)-t)}\Sigma\left(e^{-\mathrm{K}\,\mathrm{diag}(\mathbf{u}(t_i)-t)}\right)'\mathrm{d}t\right| - \frac{1}{2}\sum_{i=2}^{n}\mathrm{tr}\left\{\left(\int_{t_{i-1}}^{t_i} e^{-\mathrm{K}\,\mathrm{diag}(\mathbf{u}(t_i)-t)}\Sigma\left(e^{-\mathrm{K}\,\mathrm{diag}(\mathbf{u}(t_i)-t)}\right)'\mathrm{d}t\right)^{-1}\Delta Y_i\,\Delta Y_i'\right\}. \tag{3.6.6}$$

In this context however, differently from the case with no mean-reversion, there is no easy way to find explicitly the matrix $\hat{\Sigma}$ which maximizes the likelihood function (3.6.6). Therefore we derive our estimator based on the following moment condition:

$$\Sigma = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=2}^{n}\left\{\left(\Delta Y_i\,\Delta Y_i'\right)\oslash\mathbf{H}_i\right\}\right], \tag{3.6.7}$$

where $\mathbf{H}_i = \int_{t_{i-1}}^{t_i} e^{-\mathrm{K}\,\mathrm{diag}(\mathbf{u}(t_i)-t)}\mathbf{1}\mathbf{1}'\left(e^{-\mathrm{K}\,\mathrm{diag}(\mathbf{u}(t_i)-t)}\right)'\mathrm{d}t$.

Similarly to what we did in Chapter 2, we replace each missing value of $Y_{t_i}^{(l)}$ with its conditional expectation with respect to $\left(Y_{t_j}^{(l),Obs}\right)_{j=1,\dots,N(l)}$, i.e. the available observations of the component $Y^{(l)}$ seen in isolation. As in the Brownian motion case, conditioning with respect to the univariate time series, rather than with respect to the full set of observations (as prescribed by Expectation-Maximization algorithms - cf. Dempster et al. [1977]), has the advantage, as we will see below, of producing an estimator that has a high level of efficiency and is also simple and fast to compute given that it does not require complex matrix operations and that it can be computed in a single iteration.

The process Y is equivalent in distribution to the following time-scaled Brownian motion

$$Y_t^{(l)} \overset{d}{=} s_l\, W_{\int_{t_0^{(l)}}^{t} e^{-2\kappa_l\left(u_l(v)-v\right)}\mathrm{d}v},$$

therefore we can easily compute the required expectation by applying Brownian

bridge. Specifically we have

$$\mathbb{E}\left[Y_t^{(l)} \;\middle|\; Y_{l,d_l(t)}^{Obs}, Y_{l,u_l(t)}^{Obs}\right] = Y_{l,d_l(t)}^{Obs} + \left(Y_{l,u_l(t)}^{Obs} - Y_{l,d_l(t)}^{Obs}\right) \frac{\int_{d_l(t)}^{t} e^{-2\kappa_l\,(u_l(v)-v)}\,\mathrm{d}v}{\int_{d_l(t)}^{u_l(t)} e^{-2\kappa_l\,(u_l(v)-v)}\,\mathrm{d}v}\,.$$

Therefore we define

$$\begin{aligned}
\Delta\hat{Y}_i &:= \left(\mathbb{E}\left[\Delta Y_i^{(1)}\middle| Y_{1,:}^{Obs};\Sigma\right], \mathbb{E}\left[\Delta Y_i^{(2)}\middle| Y_{2,:}^{Obs};\Sigma\right], \ldots, \mathbb{E}\left[\Delta Y_i^{(d)}\middle| Y_{d,:}^{Obs};\Sigma\right]\right) \\
&= \Bigg(\left(Y_{1,u_1(t_i)}^{Obs} - Y_{1,d_1(t_{i-1})}^{Obs}\right) \frac{\int_{t_{i-1}}^{t_i} e^{-2\kappa_1(u_1(v)-v)}\mathrm{d}v}{\int_{d_1(t_{i-1})}^{u_1(t_i)} e^{-2\kappa_1(u_1(v)-v)}\mathrm{d}v}, \\
&\quad \left(Y_{2,u_2(t_i)}^{Obs} - Y_{2,d_2(t_{i-1})}^{Obs}\right) \frac{\int_{t_{i-1}}^{t_i} e^{-2\kappa_2(u_2(v)-v)}\mathrm{d}v}{\int_{d_2(t_{i-1})}^{u_2(t_i)} e^{-2\kappa_2(u_2(v)-v)}\mathrm{d}v}, \\
&\quad \ldots, \left(Y_{d,u_2(t_i)}^{Obs} - Y_{d,d_d(t_{i-1})}^{Obs}\right) \frac{\int_{t_{i-1}}^{t_i} e^{-2\kappa_d(u_d(v)-v)}\mathrm{d}v}{\int_{d_d(t_{i-1})}^{u_d(t_i)} e^{-2\kappa_d(u_d(v)-v)}\mathrm{d}v} \Bigg)\,.
\end{aligned}$$

We can now re-write the moment condition (3.6.7) using $\hat{Y}_i$ as predictor for the missing values, including the estimate of the prediction error calculated conditionally on the available observations and the, yet unknown, covariance Σ. We thereby obtain an iterative estimator based on the following update equation from Σ^k to Σ^{k+1}:

$$\hat{\Sigma}^{k+1} = \frac{1}{n-2}\sum_{i=2}^{n}\left[\left(\Delta\hat{Y}_i\,\Delta\hat{Y}_i{}' + \mathbb{E}\left[\Delta Y_i\,\Delta Y_i{}' - \Delta\hat{Y}_i\,\Delta\hat{Y}_i{}'\middle|\,\Pi;\hat{\Sigma}^k\right]\right) \oslash \mathbf{H}_i\right]\,. \tag{3.6.8}$$

Given that (3.6.8) computes every element $\hat{\sigma}_{l,m}$ of $\hat{\Sigma}$ (where $(\hat{\sigma}_{l,m})_{l,m=1,\ldots,d;} = \hat{\Sigma}$) independently of the others, we are able to rewrite (3.6.8) as the update equation for the pair-wise covariance $\hat{\sigma}_{l,m}$. Specifically

$$\begin{aligned}
\hat{\sigma}_{l,m}^{k+1} = &\frac{1}{n-2}\sum_{i=2}^{n}\frac{1}{\int_{t_{i-1}}^{t_i} e^{-\kappa_l(u_l(v)-v)-\kappa_m(u_m(v)-v)}\mathrm{d}v}\Big(\Delta\hat{Y}_i^{(l)}\Delta\hat{Y}_i^{(m)} \\
&+\mathbb{E}\left[\Delta Y_i^{(l)}\Delta Y_i^{(m)} - \Delta\hat{Y}_i^{(l)}\Delta\hat{Y}_i^{(m)}|\Pi;\Sigma\right]\Big)\,,
\end{aligned} \tag{3.6.9}$$

where

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{1,1} & \hat{\sigma}_{1,2} & \cdots & \hat{\sigma}_{1,d} \\ \hat{\sigma}_{2,1} & \hat{\sigma}_{2,2} & \cdots & \hat{\sigma}_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{d,1} & \hat{\sigma}_{d,2} & \cdots & \hat{\sigma}_{d,d} \end{pmatrix} .$$

In order to simplify subsequent calculations, we introduce the following measure definitions:

$$|\cdot|_a := |(r,s]|_a = \int_r^s a_t \, \mathrm{d}t, \quad |\cdot|_b := |(r,s]|_b = \int_r^s b_t \, \mathrm{d}t,$$

$$|\cdot|_c := |(r,s]|_c = \int_r^s a_t b_t \, \mathrm{d}t, \quad |\cdot|_d := |(r,s]|_d = \int_r^s a_t^2 \, \mathrm{d}t, \quad |\cdot|_o := |(r,s]|_o = \int_r^s b_t^2 \, \mathrm{d}t;$$

where

$$a_t := e^{-\kappa_l(u_l(t)-t)} \mathbb{I}_{I^l}(t), \quad b_t := e^{-\kappa_m(u_m(t)-t)} \mathbb{I}_{I^m}(t)$$
$$c_t := a_t b_t, \quad d_t := a_t^2, \quad o_t := b_t^2 .$$

We also define

$$\Delta Y_i^{(l)} := s_l \int_{t_{i-1}^{(l)}}^{t_i^{(l)}} e^{-\kappa_l(u_l(t)-t)} \, \mathrm{d}W_t^{(l)} \qquad l = 1, \ldots, d ,$$

which is the observed increment of $Y^{(l)}$ over the interval $\left(t_{i-1}^{(l)}, t_i^{(l)}\right]$.

Based on the new definitions we can re-write the update equation as

$$\hat{\sigma}_{l,m}^{k+1} = \frac{1}{n-2} \sum_{i=2}^{n} \frac{1}{|\Delta_i|_c} \left(\Delta \hat{Y}_i^{(l)} \Delta \hat{Y}_i^{(m)} + \mathbb{E}\left[\Delta Y_i^{(l)} \Delta Y_i^{(m)} - \Delta \hat{Y}_i^{(l)} \Delta \hat{Y}_i^{(m)} | \Pi; \Sigma \right] \right) ;$$

furthermore

$$\begin{aligned}
\hat{\sigma}_{l,m}^{k+1} =& \frac{1}{n-1}\sum_{i=2}^{n}\left\{\left(Y_{u_l(t_i)}^{(l)} - Y_{d_l(t_{i-1})}^{(l)}\right)\left(Y_{u_m(t_i)}^{(m)} - Y_{d_m(t_{i-1})}^{(m)}\right)\right. \\
& \cdot\frac{|\Delta_i|_d}{|\left(d_l\left(t_{i-1}\right), u_l\left(t_i\right)\right)|_d}\frac{|\Delta_i|_o}{|\left(d_m\left(t_{i-1}\right), u_m\left(t_i\right)\right)|_o}\frac{1}{|\Delta_i|_c} \\
& \left.+\hat{\sigma}_{l,m}^{k}\left(1 - \frac{|\Delta_i|_d}{|I_i^l|_d}\frac{|\Delta_i|_o}{|I_j^m|_o}\right)\right\} \\
=& \frac{1}{n-1}\sum_{i,j}\left\{\Delta Y_i^{(l)}\Delta Y_j^{(m)}\frac{|I_i^l \cap I_j^m|_d}{|I_i^l|_d}\frac{|I_i^l \cap I_j^m|_o}{|I_j^m|_o}\frac{1}{|I_i^l \cap I_j^m|_c}\right. \\
& \left.-\hat{\sigma}_{l,m}^{k}\frac{|I_i^l \cap I_j^m|_d}{|I_i^l|_d}\frac{|I_i^l \cap I_j^m|_o}{|I_j^m|_o}\right\} + \hat{\sigma}_{l,m}^{k} \\
=& \frac{1}{n-1}\sum_{i,j}\left\{\Delta Y_i^{(l)}\Delta Y_j^{(m)}\, w_{ij}^{(l,m)} - \hat{\sigma}_{l,m}^{k}\,|I_i^l \cap I_j^m|_c\; w_{ij}^{(l,m)}\right\} + \hat{\sigma}_{l,m}^{k} \\
=& \frac{1}{n-1}\left\{\left(\sum_{i,j}\Delta Y_i^{(l)}\Delta Y_j^{(m)}\, w_{ij}^{(l,m)}\right) - \hat{\sigma}_{l,m}^{k}\,\phi^{(l,m)}\right\} + \hat{\sigma}_{l,m}^{k}\ ,
\end{aligned}$$

where

$$\begin{aligned}
w_{ij}^{(l,m)} &:= \frac{|I_i^l \cap I_j^m|_d}{|I_i^l|_d}\frac{|I_i^l \cap I_j^m|_o}{|I_j^m|_o}\frac{1}{|I_i^l \cap I_j^m|_c}\ , \\
\phi^{(l,m)} &:= \sum_{i,j}|I_i^l \cap I_j^m|_c\; w_{ij}^{(l,m)}\ ,
\end{aligned}$$

and where

$$d_l(t) := \sup\left\{t_j^{(l)} :\quad t_j^{(l)} \le t\ ,\ j = 0, \ldots, N(l)\right\}\ .$$

If we now assume convergence of the algorithm, we can set $\hat{\sigma}_{l,m}^{k+1} = \hat{\sigma}_{l,m}^{k} = \hat{\sigma}_{l,m}$. From which it follows

$$\begin{aligned}
\frac{1}{n-1}\left\{\sum_{i,j}\Delta Y_i^{(l)}\Delta Y_j^{(m)}\, w_{ij}^{(l,m)} - \hat{\sigma}_{l,m}\,\phi^{(l,m)}\right\} &= 0 \\
\sum_{i,j}\Delta Y_i^{(l)}\Delta Y_j^{(m)}\, w_{ij}^{(l,m)} - \hat{\sigma}_{l,m}\,\phi^{(l,m)} &= 0\ .
\end{aligned}$$

Thus

$$\hat{\sigma}_{l,m} = \frac{1}{\phi^{(l,m)}} \sum_{i,j} \Delta Y_i^{(l)} \Delta Y_j^{(m)} \, w_{ij}^{(l,m)} = \mathrm{LW}_{l,k} \left(X_{1,\cdot}^{Obs}, X_{2,\cdot}^{Obs} \right) \ .$$

We can now write the generalization of the LW estimator for OU processes of type given by (3.6.1), including the bias correction adjustment η that needs to be applied in case the drifts of the two processes, μ_1 and μ_2, are estimated in sample.

Definition 3.6.1 (LW covariance estimator (general formula))**.** The LW estimator, $\mathrm{LW}\left(X^{(1)}, X^{(2)}\right)$, for the covariance of Ornstein-Uhlenbeck processes $X^{(1)}$ and $X^{(2)}$ (where mean-reversion is allowed only component-wise) based on non-synchronous observations $\left(X_i^{(1)}\right)_{i=1,..,N}$ and $\left(X_j^{(2)}\right)_{j=1,..,M}$ and OU-parameters a_1, a_2, κ_1, κ_2 is given by

$$\mathrm{LW}\left(X^{(1)}, X^{(2)}\right) := \begin{cases} \frac{1}{\phi} \sum_{i,j} \Delta Y_i^{(1)} \Delta Y_j^{(2)} w_{ij}, & \text{No drift} \\ \frac{1}{\phi - \eta} \sum_{i,j} \left(\Delta Y_i^{(1)} - \left| I_i^{(1)} \right|_a \overline{\mu}_1 \right) \left(\Delta Y_j^{(2)} - \left| I_j^{(2)} \right|_b \overline{\mu}_2 \right) w_{ij}, & \text{Constant drift} \end{cases}$$

where the increments are calculated as

$$\Delta Y_i^{(l)} := X_{t_i^{(l)}}^{(l)} - e^{-\kappa_l \left(t_i^{(l)} - t_{i-1}^{(l)} \right)} X_{t_{i-1}^{(l)}}^{(l)} - a_l \left(1 - e^{-\kappa_l \left(t_i^{(l)} - t_{i-1}^{(l)} \right)} \right) \qquad l = 1, 2 \ ,$$

while the weights of the estimator are given by

$$w_{ij} := \frac{\left| I_i^{(1)} \cap I_j^{(2)} \right|_d}{\left| I_i^{(1)} \right|_d} \frac{\left| I_i^{(1)} \cap I_j^{(2)} \right|_o}{\left| I_j^{(2)} \right|_o} \frac{1}{\left| I_i^{(1)} \cap I_j^{(2)} \right|_c}, \qquad \phi := \sum_{i,j} w_{ij} \left| I_i^{(1)} \cap I_j^{(2)} \right|_c ,$$

$$\eta := \frac{1}{\psi} \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_c^{\gamma} + \frac{1}{\varphi} \sum_{i,j} w_{ij} \left| I_j^{(2)} \right|_b \left| I_i^{(1)} \right|_c^{\delta} - \frac{|I \cap J|_c^{\gamma,\delta}}{\psi \ \varphi} \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b \ ,$$

The formula for the case with drift assumes that the estimates $\bar{\mu}_1$ and $\bar{\mu}_2$ are given by the *MLE* estimator, that is

$$\bar{\mu}_1 := \frac{1}{\psi} \sum_i \gamma_i \Delta Y_i^{(1)}, \quad \psi := \sum_i \gamma_i \left| I_i^{(1)} \right|_a , \quad \gamma_t := \frac{\left| I_{(t)} \right|_a}{\left| I_{(t)} \right|_d}, \quad I_{(t)} := \left\{ I_i^{(1)} : \ t \in I_i^{(1)} \right\};$$

$$\bar{\mu}_2 = \frac{1}{\varphi}\sum_j \delta_j \Delta Y_j^{(2)} \quad \varphi = \sum_j \delta_j \left| I_j^{(2)} \right|_b , \quad \delta_t := \frac{\left|J_{(t)}\right|_b}{\left|J_{(t)}\right|_o} \quad , J_{(t)} := \left\{ I_j^{(2)} : \ t \in I_j^{(2)} \right\};$$

and where

$$|\cdot|_c^\gamma : \ |(r,s]|_c^\gamma = \int_r^s c_s^{(\gamma)} \, \mathrm{d}t, \quad |\cdot|_c^\delta : \ |(r,s]|_c^\delta = \int_r^s c_s^{(\delta)} \, \mathrm{d}t, \quad |\cdot|_c^{\gamma,\delta} : \ |(r,s]|_c = \int_r^s c_s^{(\gamma,\delta)} \, \mathrm{d}t,$$

with

$$c_t^{(\gamma)} := c_t \gamma_t \, , \quad c_t^{(\delta)} := c_t \delta_t \, , \quad c_t^{(\gamma,\delta)} := c_t \gamma_t \delta_t \, ,$$
$$u_l(t) := \inf \left\{ t_i^{(l)} : \ \ t_i^{(l)} \geq t \ , \ i = 0, \ldots, N(l) \right\} \ .$$

Lemma 3.6.2. *If $X^{(1)}$ and $X^{(2)}$ are defined by stochastic differential equation (3.6.1) and $\mu \in \mathbb{R}$, then* LW $\left(X^{(1)}, X^{(2)}\right)$ *is unbiased.*

Proof. The proof of the unbiasedness of the LW estimator generalized to Ornstein-Uhlenbeck processes is given in Appendix B.1. □

3.7 Estimation with Microstructure Noise

We now consider the impact of microstructure noise on the LW estimator. Let $Y_i^{(l)} := X_i^{(l)} + u_i^{(l)}$ be an observation of process $X^{(l)}$, $l = 1, \ldots, d$, contaminated by the microstructure noise $u^{(l)}$, from which it follows that the (noisy) observed increments will be given by $\Delta Y_i^{(l)} := \Delta X_i^{(l)} + u_i^{(l)} - u_{i-1}^{(l)}$. In line with the recent literature (see Xiu [2010], Griffin and Oomen [2011], Shephard and Xiu [2012]), we work under the following assumption.

Assumption 1 (I.I.D. Gaussian Noise)**.** The noise process $u^{(l)}$ is independently and identically normally distributed, and independent of the price process and the noise processes of the other assets, with zero mean and variance $\xi_{(l)}^2$.

Calculating LW $\left(Y_{\Pi_l}^{(l)}, Y_{\Pi_k}^{(k)}\right)$ based on noise contaminated processes will still result in an unbiased estimate as long as $l \neq k$. The reason for this is that Assumption 1 ensures that $\mathbb{E}\left[\Delta X_i^{(l)} \Delta X_j^{(k)}\right] = \mathbb{E}\left[\Delta Y_i^{(l)} \Delta Y_j^{(k)}\right]$. However, the

presence of noise introduces bias in the variance estimates LW $\left(Y^{(l)}_{\Pi_l}, Y^{(l)}_{\Pi_l}\right)$ and reduces the relative efficiency of the estimator. In this section we will introduce generalized versions of the LW estimators of covariance and correlation which allow for the presence of market microstructure noise of the kind considered in Assumption 1. To this end we make the following additional assumption.

Assumption 2 (Univariate parameters known)**.** Estimates of the volatility parameters s_l of $X^{(l)}$ (as in (3.6.2)), and those of the variance of the noise term $\xi^2_{(l)}$, $l = 1, \ldots, d$, are available.

Note that Assumption 2 can be satisfied by applying one of the existing volatility estimators for univariate time series such as the QML approach of Xiu [2010].

In order to deal with the presence of microstructure noise, we observe that $\Delta Y^{(l)}_i = Y^{(l)}_{t^{(l)}_i} - Y^{(l)}_{t^{(l)}_{i-1}} = u^{(l)}_i + X^{(l)}_{t^{(l)}_i} - X^{(l)}_{t^{(l)}_{i-1}} - u^{(l)}_{i-1}$ is equivalent in distribution to $X^{(l)}_{t^{(l)}_i + \frac{\xi^2_{(l)}}{s_l}} - X^{(l)}_{t^{(l)}_{i-1} - \frac{\xi^2_{(l)}}{s_l}}$; therefore we can approach the issue of microstructure noise as if the noisy observations $Y^{(l)}_i$ were, in fact, non-synchronous observations on process $X^{(l)}$; in other words: we would like to know the increment of $X^{(l)}$ between $t^{(l)}_{i-1}$ and $t^{(l)}_i$, but it is as if the only observations we have are at $t^{(l)}_{i-1} - \frac{\xi^2_{(l)}}{s_l}$ and $t^{(l)}_i + \frac{\xi^2_{(l)}}{s_l}$. In light of this, we can adopt the same QMLE approach used is in Chapter 2, section 2.6, and in section 3.6 of this chapter. In particular, we have that the increment of $\Delta\hat{X}^{(l)}_{(t_1,t_2]} = \hat{X}^{(l)}_{t_2} - \hat{X}^{(l)}_{t_1}$, where $t^{(l)}_{i-1} \leq t_1 < t_2 \leq t^{(l)}_i$, obtained from the 'linear interpolation' of the available observations $Y^{(l)}_{t^{(l)}_{i-1}}$ $(\stackrel{d}{=} X^{(l)}_{t^{(l)}_{i-1} - \frac{\xi^2_{(l)}}{s_l}})$ and $Y^{(l)}_{t^{(l)}_i}$ $(\stackrel{d}{=} X^{(l)}_{t^{(l)}_i + \frac{\xi^2_{(l)}}{s_l}})$ is given by

$$\Delta\hat{X}^{(l)}_{(t_1,t_2]} = \Delta Y^{(l)}_i \frac{t_2 - t_1}{\left|I^{(l)}_i\right| + 2\frac{\xi^2_{(l)}}{s^2_l}} \,.$$

Based on this, following the same reasoning as in Chapter 2 (section 2.6), we can obtain the parameters for the LW estimator below:

$$\widehat{w}^{(l,k)}_{i,j} := \frac{\left|I^{(l)}_i \cap I^{(k)}_j\right|}{\left(\left|I^{(l)}_i\right| + 2\frac{\xi^2_{(l)}}{s^2_l}\right)\left(\left|I^{(k)}_j\right| + 2\frac{\xi^2_{(k)}}{s^2_k}\right)} \,,$$

and

$$\widehat{\phi}^{(l,k)} := \sum_{i,j} \left| I_i^{(l)} \cap I_j^{(k)} \right| \widehat{w}_{i,j}^{(l,k)} \ .$$

In addition, taking advantage of the relationship between parameter ϕ and the variance of the estimator described in section 3.4.2, we introduce the following data pre-processor whose aim is to find a sub-sampling of the dataset that maximizes the value of $\widehat{\phi}^{(l,k)}$ (thereby minimizing the variance of the estimate) based on the ratios between the variance of the noise $\xi^2_{(\cdot)}$ and that of the 'signal', $s^2_{\cdot}$, of the two time series, which will allow us to significantly increase the efficiency of the estimator.

Pre-processor (B)**.** Let $\Pi_{l,k} := \left(t_i^{(l,k)}\right)_{i=1,\dots,N^{(l,k)}}$ be the ordered union of the observation times in Π_l and Π_k and let

$$a\left(t_1, t_2\right) := \frac{\left(t_2 - t_1\right)^2}{\left(u_l(t_2) - d_l(t_1) + 2\frac{\xi^2_{(l)}}{s^2_l}\right)\left(u_k(t_2) - d_k(t_1) + 2\frac{\xi^2_{(k)}}{s^2_k}\right)} \ ,$$

where $u_{\cdot}(\cdot)$, $d_{\cdot}(\cdot)$ are defined as in section 3.6, and

$$c\left(t_i^{(l,k)}, t_j^{(l,k)}\right) := \begin{cases} 0 & a\left(t_i^{(l,k)}, t_j^{(l,k)}\right) + a\left(t_j^{(l,k)}, t_{j+1}^{(l,k)}\right) < a\left(t_i^{(l,k)}, t_{j+1}^{(l,k)}\right) \\ 1 & a\left(t_i^{(l,k)}, t_j^{(l,k)}\right) + a\left(t_j^{(l,k)}, t_{j+1}^{(l,k)}\right) \geq a\left(t_i^{(l,k)}, t_{j+1}^{(l,k)}\right) \end{cases} \ .$$

Starting with $i = 1$,

1. set j equal to $i + 1$,
2. calculate $c\left(t_i^{(l,k)}, t_j^{(l,k)}\right)$,
3. if $c\left(t_i^{(l,k)}, t_j^{(l,k)}\right) = 0$, discard observation $t_j^{(l,k)}$, increment index j by one unit and go to step 2,
4. if $c\left(t_i^{(l,k)}, t_j^{(l,k)}\right) = 1$, set i equal to j and go to step 1,
5. Repeat steps above until the end of the sequence.

We can now define the generalized LW estimators for noise contaminated observations.

Definition 3.7.1 (LW covariance estimator with microstructure noise)**.** Let $\widehat{\Pi}_l^{(l,k)}$ and $\widehat{\Pi}_k^{(l,k)}$ be a set of observation times resulting from applying Pre-processor A and B to Π_l and Π_k, where $l \neq k$ based on estimates s_l, s_k and $\xi^2_{(l)}, \xi^2_{(k)}$. Then the LW covariance estimator is given by

$$\widehat{\mathrm{LW}}_{l,k} \equiv \mathrm{LW}\left(X^{(l)}_{\widehat{\Pi}_l^{(l,k)}}, X^{(k)}_{\widehat{\Pi}_k^{(l,k)}}\right) := \frac{1}{\widehat{\phi}^{(l,k)}} \sum_{i,j} \Delta X_i^{(l)} \Delta X_j^{(k)} \widehat{w}_{i,j}^{(l,k)} \ , \tag{3.7.1}$$

where

$$\widehat{w}_{i,j}^{(l,k)} := \frac{\left|I_i^{(l)} \cap I_j^{(k)}\right|}{\left(\left|I_i^{(l)}\right| + 2\frac{\xi^2_{(l)}}{s_l^2}\right)\left(\left|I_j^{(k)}\right| + 2\frac{\xi^2_{(k)}}{s_k^2}\right)} \ ,$$

and

$$\widehat{\phi}^{(l,k)} := \sum_{i,j} \left|I_i^{(l)} \cap I_j^{(k)}\right| \widehat{w}_{i,j}^{(l,k)} \ .$$

Notice that when $\xi^2_{(l)} = \xi^2_{(k)} = 0$ we fall back to the no-noise case because Pre-processor A and B have no impact on the value of the estimator (i.e. we have $\mathrm{LW}\big(X^{(l)}_{\Pi_l}, X^{(k)}_{\Pi_k}\big) = \mathrm{LW}\big(X^{(l)}_{\widehat{\Pi}_l^{(l,k)}}, X^{(k)}_{\widehat{\Pi}_k^{(l,k)}}\big)$). We now give the definition of the correlation estimator for the case of observations affected by microstructure noise.

Definition 3.7.2 (LW correlation estimator with microstructure noise)**.** Let $\widehat{\Pi}_l^{(l,k)}$ and $\widehat{\Pi}_k^{(l,k)}$ be a set of observation times resulting from applying Pre-processor A and B to Π_l and Π_k, where $l \neq k$ based on estimates s_l, s_k and $\xi^2_{(l)}, \xi^2_{(k)}$. Then the LW correlation estimator is given by

$$\widehat{R}_{l,k} \equiv \widehat{R}_{l,k}\left(X^{(l)}_{\widehat{\Pi}_l^{(l,k)}}, X^{(k)}_{\widehat{\Pi}_k^{(l,k)}}\right) := \frac{\frac{1}{\widehat{\phi}^{(l,k)}} \sum_{i,j} \Delta X_i^{(l)} \Delta X_j^{(k)} \widehat{w}_{i,j}^{(l,k)}}{\sqrt{\frac{1}{\widehat{\phi}^{(l)}} \sum_{i,j} \left(\Delta X_i^{(l)}\right)^2 \widehat{w}_{i,j}^{(l,k),i}} \sqrt{\frac{1}{\widehat{\phi}^{(k)}} \sum_{i,j} \left(\Delta X_j^{(k)}\right)^2 \widehat{w}_{i,j}^{(l,k),j}}} \ , \tag{3.7.2}$$

where

$$\widehat{w}_{i,j}^{(l,k),i} := \widehat{w}_{i,j}^{(l,k)} \cdot \frac{\left|I_i^{(l)} \cap I_j^{(k)}\right|}{\left|I_i^{(l)}\right|} \ ,$$

$$\widehat{w}_{i,j}^{(l,k),j} := \widehat{w}_{i,j}^{(l,k)} \cdot \frac{\left|I_i^{(l)} \cap I_j^{(k)}\right|}{\left|I_j^{(k)}\right|} ,$$

and

$$\widehat{\phi}^{(l)} := \sum_{i,j} \left(\left|I_i^{(l)}\right| + 2\xi_{(l)}^2\right) \widehat{w}_{i,j}^{(l,k),i} ,$$

$$\widehat{\phi}^{(k)} := \sum_{i,j} \left(\left|I_j^{(k)}\right| + 2\xi_{(k)}^2\right) \widehat{w}_{i,j}^{(l,k),j} .$$

Thanks to Definition 3.7.2 now we can, similarly to what we did in section 3.4.1, give the definition of a 2-step LW estimator for the case of noise contaminated observations.

Definition 3.7.3 (LW 2-step estimator with microstructure noise)**.**

$$\mathrm{LW}_{l,k}^{2-\mathrm{step}} := \widehat{R}_{l,k}\, s_l\, s_k .$$

3.8 Simulation Results

In this section we show simulation results from applying the LW estimators introduced in this chapter. Table 3.2 compares the performance of the LW and LW 2-step estimators with that of the estimator of Hayashi and Yoshida [2005] (HY) and of the maximum likelihood (ML) estimator (computed numerically). The results have been calculated by simulating 100,000 sample paths, over the interval $[0, 1]$, of two diffusion processes, $X^{(1)}$ and $X^{(2)}$, with initial value $X_0^{(1)} = 0$ and $X_0^{(2)} = 0$, drift $\mu_{(1)} = \mu_{(2)} = 0$ and volatility parameters $\sigma_{(1)}^2 = 1$ and $\sigma_{(2)}^2 = 1$. At the start of each simulation we drew the correlation parameter $\rho_{1,2}$ from a uniform distribution $U(-1, 1)$ and generated the observation times of the two processes by applying a Poisson random sampling scheme, whereby the set of observation times is represented by the jump times of two independent Poisson processes with intensity $\lambda_{(1)} = 55$ and $\lambda_{(2)} = 55$ respectively over the same simulation interval $[0, 1]$.

The results show that both LW estimators significantly outperform HY and also that the RMSE of the newly introduced LW 2-step estimator is close to the one of ML. Intuitively, to understand the reason why LW 2-step outperforms LW,

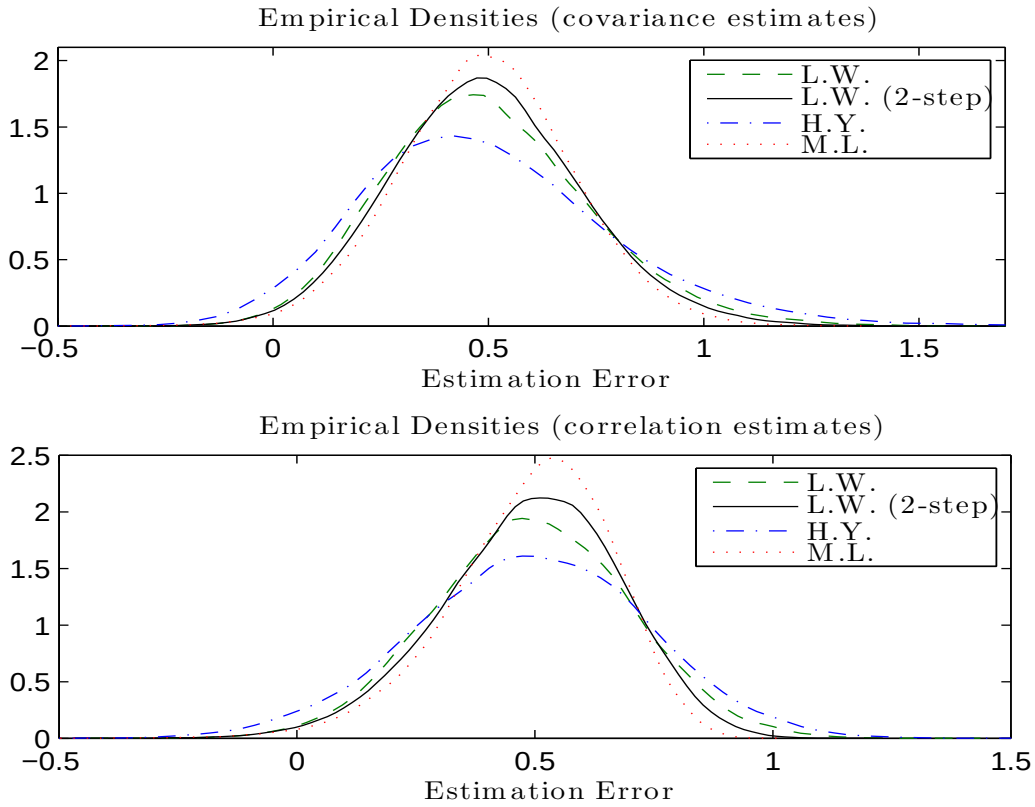


Figure 3.5: Empirical densities (by kernel density estimation) of the covariance and correlation estimates obtained using LW, LW 2-step, HY and ML (numerically), using Poisson observations and 50,000 iterations. Simulation parameters: $\sigma^2_{(1)} = 1$, $\sigma^2_{(2)} = 1$ and $\rho_{1,2} = 0.5$.

Table 3.2: Results (Poisson observations)

	Average Error	RMSE	Relative Efficiency
H.Y.	-0.0018	0.3025	154.05%
L.W.	-0.0017	0.2413	122.85%
L.W. (2-step)	-0.0019	0.2139	108.89%
M.L.	-0.0017	0.1964	100.00%

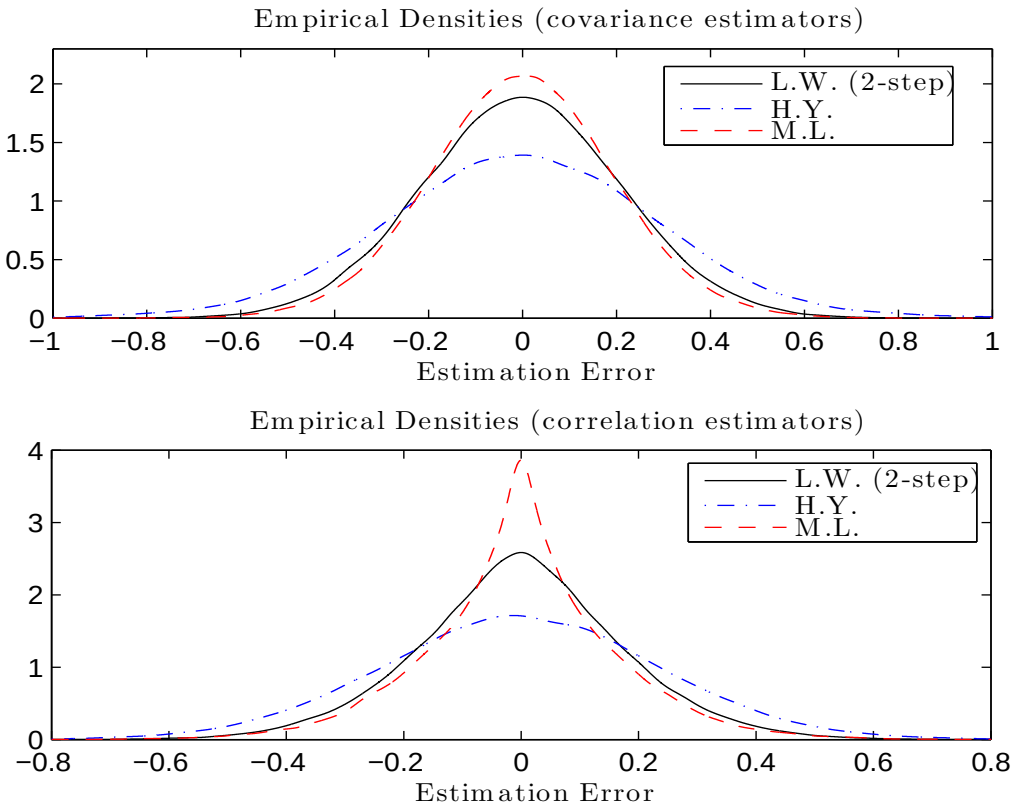


Figure 3.6: Empirical densities (by kernel density estimation) of the covariance estimates obtained using LW, LW 2-step, HY and ML (numerically), using Poisson observations and 100,000 samples. Simulation parameters: $\sigma^2_{(1)} = 1$, $\sigma^2_{(2)} = 1$ and $\rho_{1,2} \sim U(-1, 1)$.

we need to consider that both LW and HY implicitly discard from the dataset observations according to Pre-processor A, by which we mean that, given a dataset, whether we apply Pre-processor A or not, the covariance estimates of LW and HY will not change. LW 2-step avoid this waste of information by using LW specific methods only for the calculation of the correlation estimate, while estimating volatilities using the univariate LW estimator, which coincides with the efficient univariate ML estimator. Figure 3.5 displays the empirical densities by using the same simulation approach as above, except for the fact that the correlation parameter $\rho_{1,2}$ has been kept constant at each simulation with a value of 0.5, while Figure 3.6 draws random correlations from $U(-1, 1)$. Both figures show the

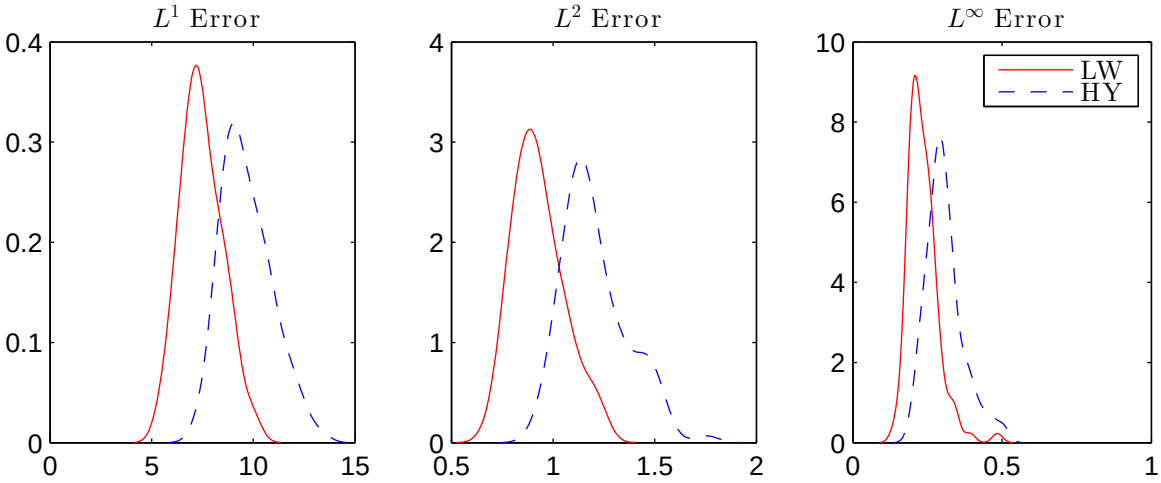


Figure 3.7: Empirical densities of the L^1, L^2 and L^∞ errors of LW and HY in the estimation of a 10×10 covariance matrix (1,000 samples) based on Poisson sampling.

empirical distribution of both covariance and correlation estimators and confirm the results of Table 3.2 with the LW estimator significantly outperforming HY with a level of accuracy that is close to that of the ML estimator. Next, Figure 3.7 compares the distributions of L^1, L^2 and L^∞ errors of HY and LW (simple) in an example where they were applied to the estimation of a 10×10 covariance matrix.

The impact of the regularization procedure of Rebonato and Jackel [2000] applied in conjunction with LW is illustrated in Figure 3.8. The plots show that the impact of regularization is very small and is noticeable mainly in the plot of L^∞ error, which highlights the fact that applying the regularization step has the effect of preventing some of the most extreme outlier estimates.

Finally, table 3.3 compares the behaviour of the estimators as we increase the level of i.i.d Gaussian microstructure noise in the samples. The results are based on 10,000 sample paths, over the time interval $[0, 1320]$, using the same simulation parameters for $X^{(1)}$ and $X^{(2)}$ as in the previous examples. Again the correlation parameter $\rho_{1,2}$ is taken from uniform distribution $U(-1, 1)$ and the observation times are based on Poisson random sampling scheme with parameters $\lambda_{(1)} = 0.0182$ and $\lambda_{(2)} = 0.0182$. It should be noted that in this example,

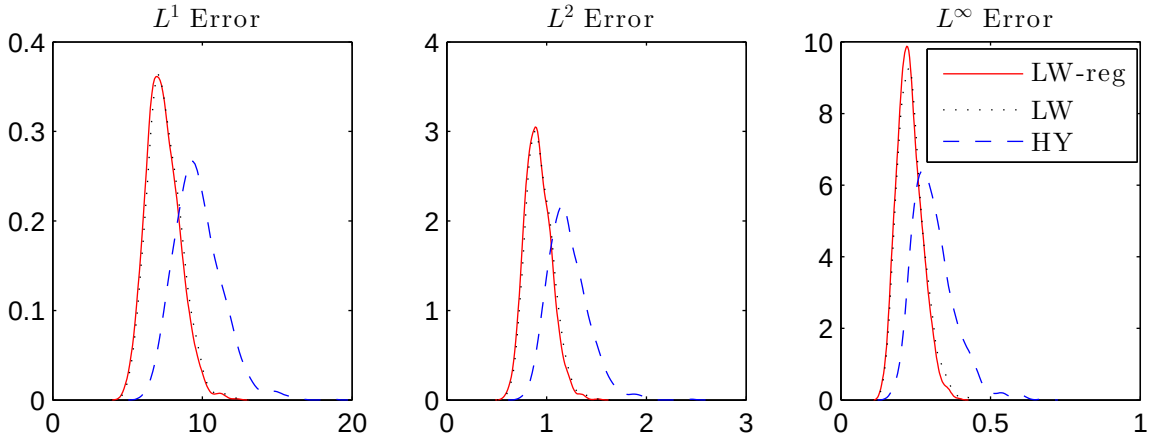


Figure 3.8: Empirical densities of the L^1, L^2 and L^∞ errors of LW and HY in the estimation of a 10×10 covariance matrix (1,000 samples) based on Poisson sampling.

based on Assumption 2, we have used the knowledge of $\sigma^2_{(1)}$, $\sigma^2_{(2)}$ and ξ in the estimation. The results show that, as we mentioned before, Pre-processor B has no impact when noise is absent. Furthermore, we can see that the simple version of LW is very sensitive to high levels of noise, however, both $\widehat{\text{LW}}$ and $\widehat{\text{LW}}$ (2-step) behave very well in the presence of noise. It is also interesting to notice that the unmodified LW (2-step) is also fairly robust to noise in the sample.

Table 3.3: Impact of microstructure noise on RMSE

ξ	HY	LW	LW (2-step)	$\widehat{\text{LW}}$	$\widehat{\text{LW}}$ (2-step)
0.0	0.3077	0.2439	0.1800	0.2439	0.1800
1.0	0.3136	0.2574	0.1917	0.2572	0.1903
2.0	0.3305	0.3041	0.2500	0.2917	0.2240
3.0	0.3604	0.3994	0.3320	0.3403	0.2689
4.0	0.4056	0.5502	0.4028	0.3990	0.3228
5.0	0.4680	0.7562	0.4559	0.4694	0.3817
6.0	0.5487	1.0155	0.4943	0.5445	0.4405
7.0	0.6480	1.3263	0.5220	0.6212	0.4995

3.9 Conclusion

In this chapter we have extended the domain of application of the Linearly Weighted (LW) estimator introduced in Chapter 2 in several ways. To begin with, we have studied how the LW estimator can be applied to the estimation of covariance matrices, showing that it is possible to take advantage of a very useful property (verified empirically) of LW, whereby the variance of the estimator is very closely linked to the value of its denominator ϕ, to derive a natural set of weights for the error measure required by the regularization procedure proposed in Rebonato and Jackel [2000]. Simulation results show that the suggested approach allows us to produce positive semi-definite covariance estimates with a RMSE (based on the Frobenius norm) that it is lower than it would be if no regularization was applied. Then, we have introduced a new correlation estimator based on LW, as well as a 2-step variant of the LW estimator of covariance which displays further efficiency improvements compared to the one introduced in Chapter 2. We have also extended the applicability of LW to the low-frequency context, by deriving the bias-corrected version of the LW estimator for the case where the drift of the process is assumed to be constant, non negligible, and estimated in sample, and also to the domain of mean-reverting Ornstein-Uhlenbeck processes, considering the case of the mean-reversion level being estimated out-of-sample as well as in sample. Finally, we have considered the potential impact of microstructure noise on the estimator and derived, under the assumption of knowledge of the univariate parameters describing the variance of noise and that of the processes under investigation, an estimation method that aims at finding a sub-sampling of the dataset that minimizes the RMSE of the estimate through the maximization of the denominator ϕ of the estimator. Such estimator is shown to be robust to the presence of noise and to consistently outperform the estimator of Hayashi and Yoshida [2005]. Concluding, simulation results tell us that the LW estimator (in its different incarnations) is able to offer a level of performance that is close to one of the maximum likelihood estimator (which is only available through computationally intensive - particularly in high dimensions - numerical procedures) while at the same time being very simple to implement and fast to compute even for high-dimensional datasets.

Part III

Asset Pricing with Counterparty Risk, Collateralization and Funding Costs

Derivative longevity risk solutions, such as bespoke and indexed longevity swaps, allow pension schemes and annuity providers to swap out longevity risk, but introduce counterparty credit risk, which can be mitigated if not fully eliminated by collateralization. We examine the impact of bilateral default risk and collateral rules on the marking to market of longevity swaps, and show how longevity swap rates must be determined endogenously from the collateral flows associated with the marking-to-market procedure. For typical interest rate and mortality parameters, we find that the impact of collateralization is modest in the presence of symmetric default risk, but more pronounced when default risk and/or collateral rules are asymmetric. Our results suggest that the overall cost of collateralization is comparable with, and often much smaller than, that found in the interest-rate swaps market, which may then provide the appropriate reference framework for the credit enhancement of both indemnity-based and indexed longevity risk solutions. The next chapter reproduces Biffis, Blake, Pitotti, and Sun [2012].

Chapter 4

The Cost of Counterparty Risk and Collateralization in Longevity Swaps[1]

4.1 Introduction

The market for longevity-linked securities and derivatives has recently experienced a surge in transactions in longevity swaps. These pure longevity hedges are agreements between two parties to exchange fixed payments against variable payments linked to the number of survivors in a reference population [see Dowd et al., 2006]. Table 4.1 presents a list of recent deals that have been publicly disclosed. So far, transactions have mainly involved pension funds and annuity providers wanting to hedge their exposure to longevity risk[2] but without having to bear any basis risk. The variable payments in such longevity swaps are designed to match precisely the mortality experience of each individual hedger: hence the name bespoke longevity swaps. This is essentially a form of longevity risk insurance, similar to annuity reinsurance in reinsurance markets. Indeed, most of the longevity swaps executed to date have been bespoke, indemnity-based swaps of the kind familiar in reinsurance markets. This is true despite the fact that some of

[1] This chapter reproduces Biffis, Blake, Pitotti, and Sun [2012].

[2] By longevity risk we mean exposure to the systematic risk of mortality improvements, which cannot be mitigated by pooling together large numbers of lives.

the swaps listed in table 4.1 have been arranged by investment banks: the banks have worked with insurance companies (in some cases insurance company subsidiaries) in order to deliver a solution in a format familiar to the counterparty. A fundamental difference from other forms of reinsurance, however, is that longevity swaps are typically collateralized, whereas typical insurance/reinsurance transactions are not.[1] The main reason is that longevity swaps are often part of a wider de-risking strategy involving other collateralized instruments (interest-rate and inflation swaps, for example), and also the fact that hedgers have been increasingly concerned with counterparty risk[2] in the wake of the Global Financial Crisis of 2008-09. In this article, we provide a framework to quantify the trade-off between the exposure to counterparty risk in longevity swaps and the cost of credit enhancement strategies such as collateralization.

As there is no accepted framework yet for marking to market/model longevity swaps, hedgers and hedge suppliers look to other markets to provide a reference model for counterparty risk assessment and mitigation. In interest-rate swap markets, for example, the most common form of credit enhancement is the posting of collateral. According to the International Swap and Derivatives Association (ISDA) almost every swap at major financial institutions is 'bilaterally' collateralized [ISDA, 2010b], meaning that either party is required to post collateral depending on whether the market value of the swap is positive or negative.[3] The vast majority of transactions is collateralized according to the Credit Support Annex (CSA) to the Master Swap Agreement introduced by ISDA [1994]. The Global Financial Crisis highlighted the importance of bilateral counterparty

[1] One rationale for this is that reinsurers aggregate several uncorrelated risks and pooling/diversification benefits compensate for the absence of collateral [e.g., Cummins and Trainar, 2009; Lakdawalla and Zanjani, 2007]. Insurers/reinsurers are still required by their regulators to post regulatory or solvency capital which plays a similar role to collateral but at aggregate level.

[2] Basel II [2006, Annex 4] defines counterparty risk as 'the risk that the counterparty to a transaction could default before the final settlement of the transaction's cash flows'. The recent Solvency II proposal makes explicit allowance for a counterparty risk module in its 'standard formula' approach; see CEIOPS [2009].

[3] 'Unlike a firm's exposure to credit risk through a loan, where the exposure to credit risk is unilateral and only the lending bank faces the risk of loss, counterparty credit risk creates a bilateral risk of loss: the market value of the transaction can be positive or negative to either counterparty to the transaction. The market value is uncertain and can vary over time with the movement of underlying market factors.' [Basel II, 2006, Annex 4].

risk and collateralization for over-the-counter markets, spurring a number of responses [e.g, Assefa et al., 2010; Brigo and Capponi, 2009; Brigo et al., 2011, 2012; ISDA, 2009]. The Dodd-Frank Wall Street Reform and Consumer Protection Act (signed into law by President Barack Obama on July 21, 2010) is likely to have a major impact on the way financial institutions will manage counterparty risk in the coming years.[1] The recently established Life and Longevity Markets Association (LLMA) [2] has counterparty risk at the center of its agenda, and will certainly draw extensively from the experience garnered in fixed-income and credit markets.

Collateralization strategies address the concerns aired by pension trustees regarding the efficacy of longevity swaps, but introduce another dimension in the traditional pricing framework used for insurance transactions. The 'insurance premium' embedded in a longevity swap rate reflects not only the aversion (if any) of the counterparties to the risk being transferred and the cost of regulatory capital involved in the transaction, but also the expected costs to be incurred from posting collateral during the life of the swap. To quantify the impact of collateral on swap rates, we must examine the sensitivity of the counterparties to the cost of this form of credit risk mitigation. Let us first take the perspective of a hedge supplier (reinsurer or investment bank) issuing a collateralized longevity swap to a counterparty (pension fund or annuity provider). Whenever the swap is sufficiently out-of-the-money to the hedge supplier, the hedge supplier is required to post collateral, which can be used by the hedger to mitigate losses in the event of default. Although interest on collateral is typically rebated, there is both a funding cost and an opportunity cost, as the posting of collateral depletes the resources the hedge supplier can use to meet her capital requirements at aggregate level as well as to write additional business. On the other hand, whenever the swap is sufficiently in-the-money to the hedge supplier, the hedge supplier will receive collateral from the counterparty, thus benefiting from capital relief in regulatory valuations and freeing up capital that can be used to sell additional longevity protection. The benefits can be far larger if collateral can be

[1] See, for example, 'Berkshire may scale back derivative sales after Dodd-Frank', *Bloomberg*, August 10, 2010.

[2] See http://www.llma.org.

re-pledged for other purposes, as in the interest-rate swaps market.[1] The same considerations can be made from the viewpoint of the hedger, but the funding needs and opportunity costs of the two parties are unlikely to offset each other exactly. This is particularly relevant for transactions involving parties subject to different regulatory frameworks. In the UK and several other countries, for example, longevity risk exposures are more capital intensive for hedge suppliers, such as insurers, than for pension funds.[2]

In the absence of collateral, and ignoring longevity risk aversion, swap rates[3] depend on best estimate survival probabilities for the hedged population and on the degree of covariation between the floating leg of the swap and the defaultable term structure of interest rates facing the hedger and the hedge supplier.[4] This means that a proper analysis of a longevity swap cannot disregard the sponsor's covenant when the hedger is a pension plan (see section 4.3 below). In the presence of collateralization, longevity swap rates are also shaped by the expected collateral costs, and swap valuation formulae involve a discount rate reflecting the cost of collateral. As a result, default-free valuation formulae are not appropriate even in the presence of full collateralization and the corresponding absence of default losses.[5]

We quantify collateral costs in two ways: i) in terms of funding costs that are incurred or mitigated when collateral is posted or received, and ii) as the opportunity cost of selling additional longevity protection. In both cases, we find that, for typical interest rate and mortality parameters, the impact of collateralization on swap rates is modest when default risk and collateral rules are symmetric. There are two main reasons for this. First, the different nature of the risk on which the swap is written, a floating rate in the case of interest-rate swaps, a

[1] According to ISDA [2010b],the vast majority of collateral is rehypothecated for other purposes in interest-rate swap markets. Currently, collateral can be re-pledged under the New York Credit Support Annex, but not under the English Credit Support Deed [see ISDA, 2010a].

[2] This asymmetry is, in part, a by-product of rules allowing, for example, pension liabilities to be quantified by using outdated mortality tables or discount rates reflecting optimistic expected returns.

[3] Defined as the rates in the fixed legs of the swap zeroing its market value at inception.

[4] Along the same lines, Inkmann and Blake [2010] show how the discount rate for the valuation of pension liabilities should reflect funding risk.

[5] See Johannes and Sundaresan [2007] for the case of symmetric default risk and full collateralization in interest-rate swaps.

smoother survival curve in the case of longevity swaps. Second, the countervailing effects of longevity risk and interest-rate risk dilute the overall impact of collateralization on swap rates:

i) On the one hand, the receiver of the fixed survival rate (the hedge supplier) posts collateral when mortality is lower and hence longevity exposures are more capital intensive. On the other hand, she receives collateral when mortality is higher and longevity protection is less capital intensive. The overall effect is to push (fixed) swap rates higher, to compensate the hedge supplier for the positive dependence between collateral posting and capital costs.

ii) When the hedger or hedge supplier is out-of-the-money, collateral outflows are larger in low interest rate environments (i.e., when liabilities are discounted at a lower rate), hence there is a negative relationship between the amount of collateral posted and the counterparties' funding/opportunity costs. This mitigates the overall impact of collateralization on longevity swap rates.

When default risk and/or collateral rules are asymmetric, the opposing effects are of different magnitudes and, as a result, the impact of collateral costs on longevity swap rates is larger. For example, we find that swap rates decrease substantially when the hedger has a lower credit standing (i.e., higher funding costs) and collateral rules are more favorable to the hedge supplier. Although collateralization introduces an explicit link between the individual risk exposures and the hedge supplier's funding risk (hence some of the pooling/diversification benefits used to substitute for collateralization in the standard insurance model may be lost), in our examples we find that the opposite effects of longevity and interest rate risk make the overall impact of collateralization comparable with, and typically lower than, that observed in fixed-income markets [e.g., Johannes and Sundaresan, 2007]. An important implication is that the interest-rate swaps market might provide an appropriate framework for the collateralization of bespoke longevity solutions, even though such solutions lack of the transparency and standardization benefits associated with indexed-based instruments. Investment banks have sold index-based longevity swaps which have a structure that would be more familiar to capital markets investors, but they have so far been less popular than bespoke solutions to date. Nevertheless, for the longevity swaps market to really take off, it is necessary to expand beyond the limits of the reinsurance

market and attract such new investors.

On the methodological side, we show how longevity swap rates must be determined endogenously from the dynamic marking to market[1] of the swap and the collateral rules specified by the contract. To see why, note that the market value of the swap at each valuation date depends on the evolution of the relevant state variables (mortality, interest rates, credit spreads), as well as on the swap rate locked in at inception. On the other hand, the swap's market value will typically affect collateral amounts and, in a setting where collateral is costly, will embed the market value of the costs associated with future collateral flows. Hence, the swap rate can only be determined by explicitly taking into account the marking-to-market process and the dynamics of collateral posting. To avoid the computational burden of nested Monte Carlo simulations, we use an iterative procedure based on the Least-Squares Monte Carlo (LSMC) approach[2] [see Glassermann, 2004, and references therein]. We provide several numerical examples showing how different collateralization rules shape longevity swap rates giving rise to margins in (best estimate) survival probabilities reflecting the cost of future collateral flows. Although our focus is on longevity risk solutions, the approach can be applied to other instruments, such as over-the-counter solutions for inflation and credit risk.

Our work contributes to the existing literature on longevity risk pricing in at least three ways: i) we introduce default risk in the pricing of longevity risk solutions, and properly address its bilateral nature; ii) we explicitly allow for collateralization rules, which are the backbone of any real-world hedging solution and materially affect the pricing of over-the-counter transactions; iii) we introduce a 'structural' dimension in an otherwise reduced-form pricing framework, by allowing for funding/opportunity costs associated with longevity risk exposures held by hedgers and hedge suppliers. As there is essentially no pub-

[1] Here and in what follows, by 'market value' and 'marking to market' we mean that assets and liabilities are valued according to a market-consistent accounting/regulatory standard.

[2] A similar approach is used by Bacinello et al. [2009, 2010] for surrender guarantees in life policies and by Bauer et al. [2010b] for the computation of capital requirements within the Solvency II framework. The term American Monte Carlo is often used in financial engineering to refer to this approach. We stick to the term Least Squares Monte Carlo, as it is more common in the insurance industry [e.g., Hörig and Leitschkis, 2012].

licly available information on swap rates, our approach[1] has the advantage of using publicly available information on credit markets and regulatory standards, without having to rely exclusively on calibration to primary insurance market prices, approximate hedging methods or assumptions on agents' risk preferences [e.g., Bauer et al., 2010a, 2012; Biffis et al., 2010; Chen and Cummins, 2010; Cox et al., 2010; Dowd et al., 2006; Ludkovski and Young, 2008, among others].

The article is organized as follows. In the next section, we introduce longevity swaps and formalize their payoffs. Although the setup covers the case of both bespoke and index-based swaps, we focus on the former to keep the paper focused. In section 4.2.1, we examine the marking to market of a longevity swap during its lifetime to demonstrate the impact of counterparty risk on the hedger's balance sheet. Section 4.3 introduces bilateral default risk in longevity swap valuation formulae. We identify the main channels through which default risk affects the market value of swaps and show why an iterative procedure is needed to compute swap rates. Section 4.4 introduces credit enhancement in the form of collateralization, and shows how longevity swap rates are affected even in the presence of full cash collateralization (and hence absence of default losses). We compute swap rates by using an iterative procedure based on the LSMC approach. In section 4.5, several stylized examples are provided to understand how different collateralization rules may affect longevity swap rates. Concluding remarks are offered in section 4.6. Further details and technical remarks are collected in an appendix.

4.2 Longevity swaps

We consider a hedger (insurer selling annuities, pension fund), referred to as party H, and a hedge supplier (reinsurer, investment bank), referred to as counterparty HS. Agent H has the obligation to pay amounts $X_{T_1}, X_{T_2}, \ldots$, possibly dependent on interest rates and inflation, to each survivor at fixed dates $0 < T_1 \leq T_2, \ldots$ of an initial population of n individuals alive at time zero (annuitants or pensioners). We are clearly restricting our attention to homogeneous liabilities for ease of

[1] Similarly, Biffis and Blake [2009, 2010a] endogenize longevity risk premia by introducing asymmetric information and capital requirements in a risk-neutral setting.

Date	Hedger	Size	Term (yrs)	Type	Interm./supplier
Jan 2008	Lucida	N.A.	10	indexed	JP Morgan ILS funds
Jul 2008	Canada Life	GBP 500m	40	bespoke	JP Morgan ILS funds
Feb 2009	Abbey Life	GBP 1.5bn	run-off	bespoke	Deutsche Bank ILS funds / Partner Re
Mar 2009	Aviva	GBP 475m	10	bespoke	Royal Bank of Scotland
Jun 2009	Babcock International	GBP 750m	50	bespoke	Credit Suisse Pacific Life Re
Jul 2009	RSA	GBP 1.9bn	run-off	bespoke	Goldman Sachs (Rothesay Life)
Dec 2009	Berkshire Council	GBP 750m	run-off	bespoke	Swiss Re
Feb 2010	BMW UK	GBP 3bn	run-off	bespoke	Deutsche Bank Paternoster
Dec 2010	Swiss Re (Kortis bond)	USD 50m	8	indexed	ILS funds
Feb 2011	Pall (UK) Pension Fund	GBP 70m	10	indexed	JP Morgan
Aug 2011	ITV	GBP 1.7bn	N.A.	bespoke	Credit Suisse
Nov 2011	Rolls Royce	GBP 3bn	N.A.	bespoke	Deutsche Bank
Dec 2011	British Airways	GBP 1.3bn	N.A.	bespoke bespoke	Goldman Sachs (Rothesay Life)
Jan 2012	Pilkington	GBP 1bn	N.A.	bespoke	Legal & General
Apr 2012	Berkshire Council	GBP 100m	run-off	bespoke	Swiss Re

Table 4.1: Publicly announced longevity swap transactions 2008-2012.

exposition, more general situations requiring obvious modifications. Party H's liability at a generic payment date $T > 0$ is given by the random variable $(n - N_T)X_T$, where N_T counts the number of deaths experienced by the population during the period $[0, T]$. Assuming that the individuals' death times have common intensity[1] $(\mu_t)_{t\geq 0}$, the expected number of survivors at time T can be written as $E^{\mathbb{P}}[n - N_T] = np_T$, with the survival probability p_T given by (see the appendix)

$$p_T := E^{\mathbb{P}}\left[\exp\left(-\int_0^T \mu_t \mathrm{d}t\right)\right]. \tag{4.2.1}$$

Here and in the following, $\mathbb{P}$ denotes the real-world probability measure. The intensity could be modeled by using, for example, any of the stochastic mortality

[1] As discussed more in detail in the appendix, for tractability we restrict our attention to the case of doubly stochastic (or Cox, conditionally Poisson) death times.

models considered in Cairns et al. [2009]. For our examples, we will rely on the simple Lee-Carter mortality projection model [Lee and Carter, 1992].

Let us now consider a financial market and introduce the risk-free rate process $(r_t)_{t\geq 0}$ (in practice, an overnight rate). We assume that a market-consistent price of the liabilities can be computed by using a risk-neutral measure $\widetilde{\mathbb{P}}$, equivalent to $\mathbb{P}$, such that the death times have the same intensity process $(\mu_t)_{t\geq 0}$ [with different dynamics, in general, under the two measures; see Biffis et al., 2010]. The time-0 market value of the aggregate liability can then be written as

$$E^{\widetilde{\mathbb{P}}}\left[\sum_i \exp\left(-\int_0^{T_i} r_t \mathrm{d}t\right)(n - N_{T_i})X_{T_i}\right] = n\sum_i E^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^{T_i}(r_t+\mu_t)\mathrm{d}t\right)X_{T_i}\right].$$

For the moment, we take the pricing measure as given: we will give it more structure later on.

We consider two instruments which H can enter into with HS to hedge its exposure: a bespoke longevity swap and an index-based longevity swap. In these swaps, in contrast with interest rate swaps, the fixed leg will be a series of fixed rates each one pertaining to an individual payment date. The reason is that mortality increases substantially at old ages and a single fixed rate would introduce a growing mismatch between the cashflows provided by the swap and those needed by the hedger. However, as with interest rate swaps, we can treat a longevity swap as a portfolio of forward contracts on the underlying floating (survival) rate.[1] In this section, we ignore default risk and focus on individual payments at maturity $T > 0$. Throughout the article, we always assume the perspective of the hedger.

A **bespoke longevity swap** allows party H to pay a fixed rate $\overline{p}^N \in (0,1)$ against the realized survival rate experienced by the population between time zero and time T. Assuming a notional amount equal to the initial population

[1] With a slight abuse of terminology, we use the term 'swap rate' for individual forward rates as well as for swap curves (a series of swap rates). We note that swap curves are often summarized by the improvement factor applied to the survival probabilities of a reference mortality table/model; see examples in section 4.5.

size, n, the net payout to the hedger at time T is[1]

$$n\left(\frac{n-N_T}{n}-\overline{p}^N\right),$$

i.e., the difference between the realized number of survivors and the pre-set number of survivors $n\overline{p}^N$ agreed at inception. Letting S_0 denote the market value of the swap at inception, we can write

$$\begin{aligned} S_0 &= nE^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T r_t\mathrm{d}t\right)\left(\frac{n-N_T}{n}-\overline{p}^N\right)\right] \\ &= nE^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T (r_t+\mu_t)\mathrm{d}t\right)\right]-nB(0,T)\overline{p}^N, \end{aligned} \tag{4.2.2}$$

with $B(0,T)$ denoting the time-zero price of a zero-coupon bond with maturity T. By setting $S_0=0$, we obtain the swap rate as

$$\overline{p}^N=\widetilde{p}_T+B(0,T)^{-1}\mathrm{Cov}^{\widetilde{\mathbb{P}}}\left(\exp\left(-\int_0^T r_t\mathrm{d}t\right),\exp\left(-\int_0^T \mu_t\mathrm{d}t\right)\right), \tag{4.2.3}$$

where the risk-adjusted survival probability $\widetilde{p}_T$ is defined as in (4.2.1) with expectations taken under $\widetilde{\mathbb{P}}$. Expression (4.2.3) shows that if the intensity of mortality is uncorrelated with bond market returns (a reasonable first-order approximation), the longevity swap curve just involves the survival probabilities $\{\widetilde{p}_{T_i}\}$ relative to the different maturities $\{T_i\}$. Several studies have recently addressed the issue of how to quantify risk-adjusted survival probabilities, for example, by calibration to annuity prices and books of life policies traded in secondary markets, or by use of approximate hedging methods (see references in Section 4.1). As there is essentially no publicly available information on swap rates, for our numerical examples we will suppose a baseline case in which $\widetilde{p}_{T_i}=p_{T_i}$ for each maturity T_i and focus on how counterparty default risk and collateral requirements might generate a positive or negative spread on best estimate survival rates. This is consistent with market practice where counterparties would agree on a real-world

[1]For ease of exposition, here and in the following sections, we consider contemporaneous settlement only. Other settlement conventions (e.g. in arrears) have negligible effects, but make valuation formulae more involved when bilateral and asymmetric default risk is introduced.

mortality model (and estimation methodology) to mark-to-model the swap at future dates. Although in what follows, we mainly concentrate on longevity risk, in practice, the floating payment of a longevity swap might involve an interbank rate component (e.g., LIBOR) or survival indexation rules different from the ones considered above. To keep the setup general, we will at times consider instruments making a generic variable payment, P, and write the corresponding swap rate $\overline{p}$ as

$$\overline{p} = E^{\widetilde{\mathbb{P}}}[P] + B(0,T)^{-1}\mathrm{Cov}^{\widetilde{\mathbb{P}}}\left(\exp\left(-\int_0^T r_t \mathrm{d}t\right), P\right). \tag{4.2.4}$$

The setup can easily accommodate **index-based longevity swaps**, standardized instruments allowing the hedger to pay a fixed rate $\overline{p}^I \in (0,1)$ against the realized value of a survival index $(I_t)_{t\geq 0}$ at time T. The latter might reflect the mortality experience of a reference population closely matching[1] that of the liability portfolio. Examples are represented by the LifeMetrics indices developed by J.P. Morgan, the Pensions Institute and Towers Watson,[2] or the Xpect indices developed by Deutsche Börse.[3] The relative advantages and disadvantages of index-based versus bespoke swaps are discussed, for example, in Biffis and Blake [2010b]. Assuming that the index admits the representation $I_t = \exp(-\int_0^t \mu_s^I \mathrm{d}s)$, with $(\mu_t^I)_{t\geq 0}$ the intensity of mortality of a reference population, the swap rate $\overline{p}^I$ is given again by expression (4.2.3), but with the process μ replaced by μ^I, and with $\widetilde{p}_T$ replaced by the corresponding risk-adjusted survival probability $\widetilde{p}_T^I$.

4.2.1 The marking-to-market (MTM) process

Longevity swaps are not currently exchange traded and there is no commonly accepted framework for counterparties to mark to market/model their positions.[4] The presence of counterparty default risk and collateralization rules, however, makes the MTM procedure a very important feature of these transactions for at least two reasons. First, at each payment date, the difference between the variable and pre-set payment generates a cash inflow or outflow to the hedger,

[1] The risk of mismatch is called basis risk. See, for example, Coughlan et al. [2011], Stevens et al. [2010b], and Salhi and Loisel [2010] for some results related to this risk dimension.

[2] See http://www.lifemetrics.com. The indices were transferred to LLMA in 2011.

[3] See http://www.xpect-index.com.

[4] At the time of writing, LLMA was working on this issue.

depending on the evolution of mortality. In the absence of basis risk (which is the case for bespoke solutions), these differences show a pure 'cashflow hedge' of the longevity exposure in operation. However, as market conditions change (e.g., mortality patterns, counterparty default risk), the impact of the swap on the hedger's balance sheet can evolve dramatically. For example, even if the swap payments are expected to provide a good hedge against longevity risk, the hedger's position will weaken considerably if the expected present value of the net payments shrinks due to deterioration in the hedge supplier's credit quality. Second, for solvency requirements, it is important to value a longevity swap under extreme market/mortality scenarios ('stress testing'). This means, for example, that even if a longevity swap qualifies as a liability on a market-consistent basis, it might still provide considerable capital relief when valued on a regulatory basis due to its recognized effectiveness as a hedge.

To illustrate some of these points, let us consider the hypothetical situation of an insurer H with a liability represented by a group of ten thousand 65-year-old annuitants drawn from the population of England & Wales in 1980. We assume that party H entered a 25-year pure longevity swap in 1980 and we follow the evolution of the contract until maturity. The population is assumed to evolve according to the death rates reported in the Human Mortality Database (HMD) for England & Wales.[1] We assume that interest-rate risk is hedged away through interest rate swaps, locking in a rate of 5% throughout the life of the swap. The role of collateral is examined later on; here, we show how the hedging instrument operates from the point of view of the hedger. For this bespoke solution, the market value of each floating-for-fixed payment occurring at a generic date T can be computed by using the valuation formula

$$S_t = nE_t^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_t^T r_s \mathrm{d}s\right)\left(\frac{n-N_t}{n}\exp\left(-\int_t^T \mu_s \mathrm{d}s\right)\right)\right] - nB(t,T)\overline{p}^N, \tag{4.2.5}$$

for each time t in $[0,T]$ at which no default has yet occurred, with $B(t,T)$ denoting the market value of a zero-coupon bond with time to maturity $T-t$, and $E_t^{\widetilde{\mathbb{P}}}[\cdot]$ the conditional expectation under a pricing measure $\widetilde{\mathbb{P}}$, given the informa-

[1] See http://www.mortality.org.

tion available at time t. As a simple benchmark case, we assume that market participants receive information from the HMD and use the Lee-Carter model to value longevity-linked cashflows. In other words, at each MTM date (including inception), longevity swap rates are based on Lee-Carter forecasts computed using the latest HMD information available.[1] Figure 4.1 illustrates the evolution of swap survival rates for an England & Wales cohort tracked from age 65 in 1980 to age 90 in 2005. It is clear that the systematic underestimation of mortality improvements by the Lee-Carter model in this particular example will mean that the hedger's position will become increasingly in-the-money as the swap matures. This is shown in figure 4.2. In practice, the contract may allow the counterparty to cancel the swap or re-set the fixed leg for a nonnegative fee, but we ignore these features in this example. Figure 4.2 also reports the sequence of net cashflows generated by the swap. As interest rate risk is hedged away − and again ignoring default risk for the moment − cash inflows/outflows arising in the backtesting exercise only reflect the difference between the realized survival rates and the swap rates locked in at inception. On the other hand, the swap's market value reflects changes in market swap rates, which by assumption follow the updated Lee-Carter forecasts plotted in figure 4.1 and differ from the realized survival rates. As is evident from figure 4.2, the credit exposure of a longevity swap is close to zero at inception and at maturity, but may be sizable in between, depending on the trade-off between changes in market/mortality conditions and the residual swap payments (amortization effect). The credit exposure is quantified by the replacement cost, i.e., the cost that the nondefaulting counterparty would have to incur at the default time to replace the instrument at market prices then available. As a simple example which predicts the next section, let us introduce credit risk (but no default) and assume that in 1988 the credit spread of the hedge supplier widens across all maturities by 50 and 100 basis points. The impact of these two scenarios on the hedger's balance sheet is dramatic, as shown in figures 4.2-4.3, demonstrating how MTM profits and losses can jeopardize a successful cashflow hedge.

[1] See Cairns et al. [2011]; Dowd et al. [2010a,b] for a comprehensive analysis of alternative mortality models; see also Girosi and King [2008] and Pitacco et al. [2009].

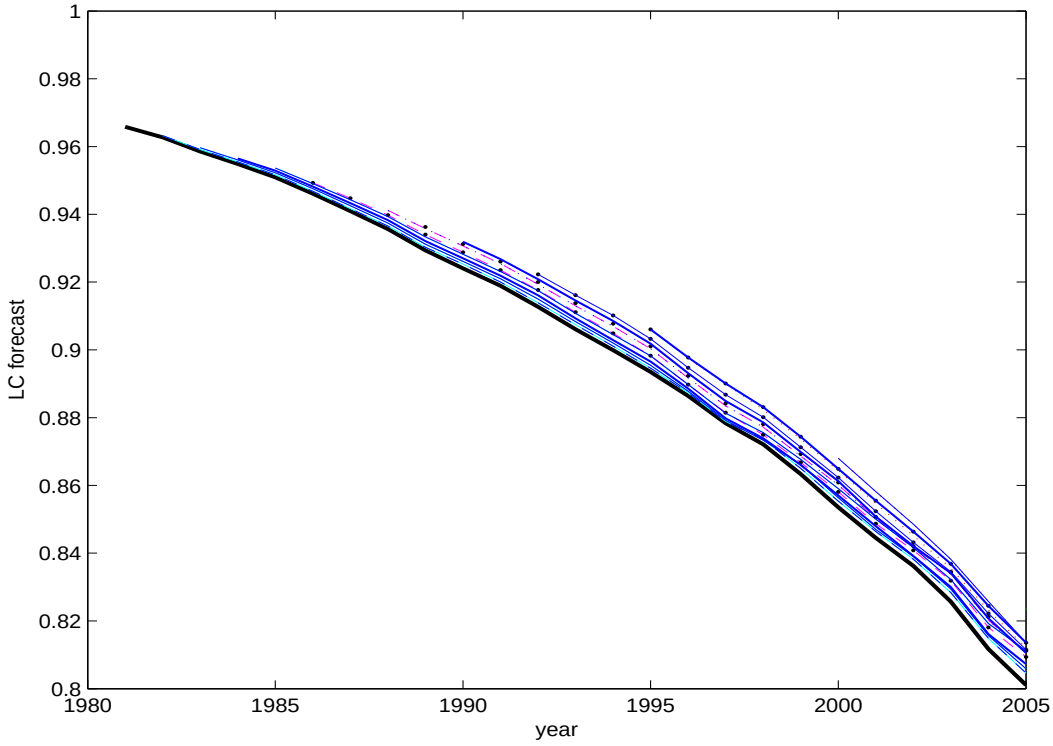


Figure 4.1: Survival curves computed at the beginning of each year $t = 1980, \ldots, 2004$ for England & Wales males aged $65 + t - 1980$ in year t. Forecasts are based on the Lee-Carter model using the latest Human Mortality Database data available at the beginning of each year t.

4.3 Counterparty default risk

The backtesting exercise of the previous section has demonstrated the importance of the hedge supplier's credit risk and the marking to market procedure in assessing the value of a longevity swap to the hedger. A correct approach, however, should allow for the fact that counterparty risk is bilateral. This is the case even when the hedger is a pension plan. Private sector defined benefit pension plans in countries such as the UK are founded on trust law and rely on a promise by (rather than a guarantee from) the sponsoring employer to pay the benefits to plan members. This promise is known as the 'sponsor covenant'. The strength of the sponsor covenant depends on both the financial strength of the employer

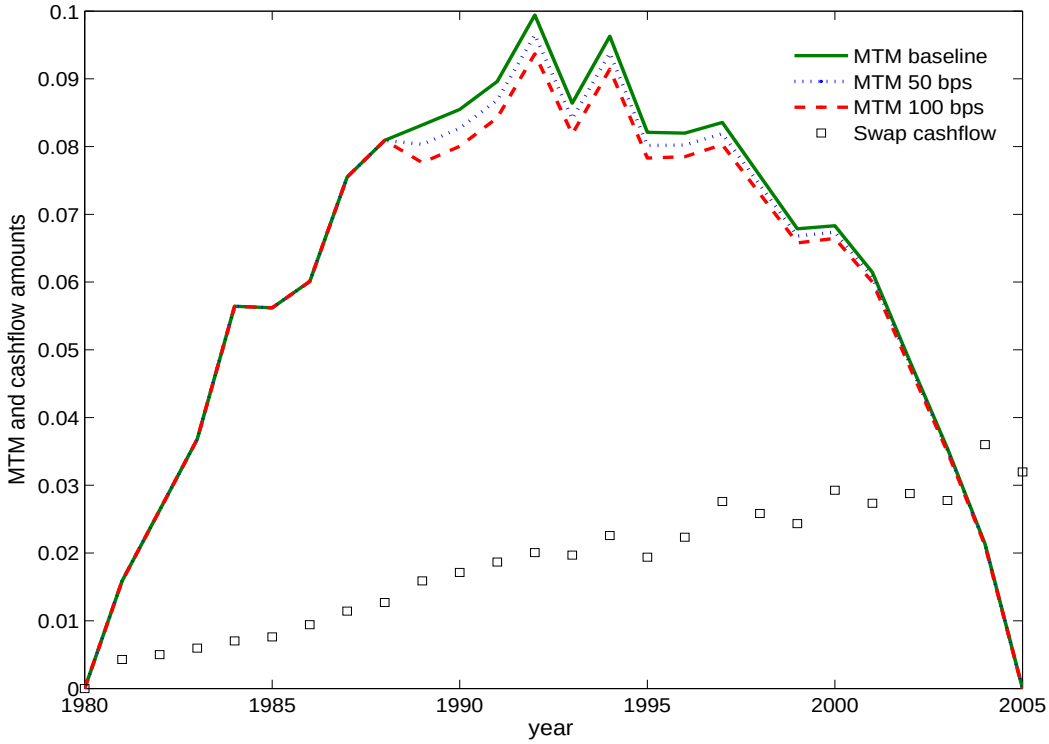


Figure 4.2: Mark-to-market value of the longevity swap in the baseline case and with counterparty B's credit spread widening by 50 and 100 basis points over 1988-2005. In the absence of default, the net payments from the swap are insensitive to credit spread changes.

and the employer's commitment to the scheme.[1] As a reasonable but imperfect proxy for the effect of the sponsor covenant, we use the sponsor's default intensity (party H's default intensity). For large corporate pension plans, the intensity can be derived/extrapolated from spreads observed in corporate bond and CDS markets. For smaller plans, an analysis of the funding level and strategy of the scheme is required.

Assume that both party H and HS may default at random times $\tau^{\mathrm{H}}, \tau^{\mathrm{HS}}$, admit-

[1]In the UK, for example, The Actuarial Profession [2005, par. 3.2] defined the sponsor covenant as: "the combination of (a) the ability and (b) the willingness of the sponsor to pay (or the ability of the trustees to require the sponsor to pay) sufficient advance contributions to ensure that the scheme's benefits can be paid as they fall due." See also The Pensions Regulator [2009].

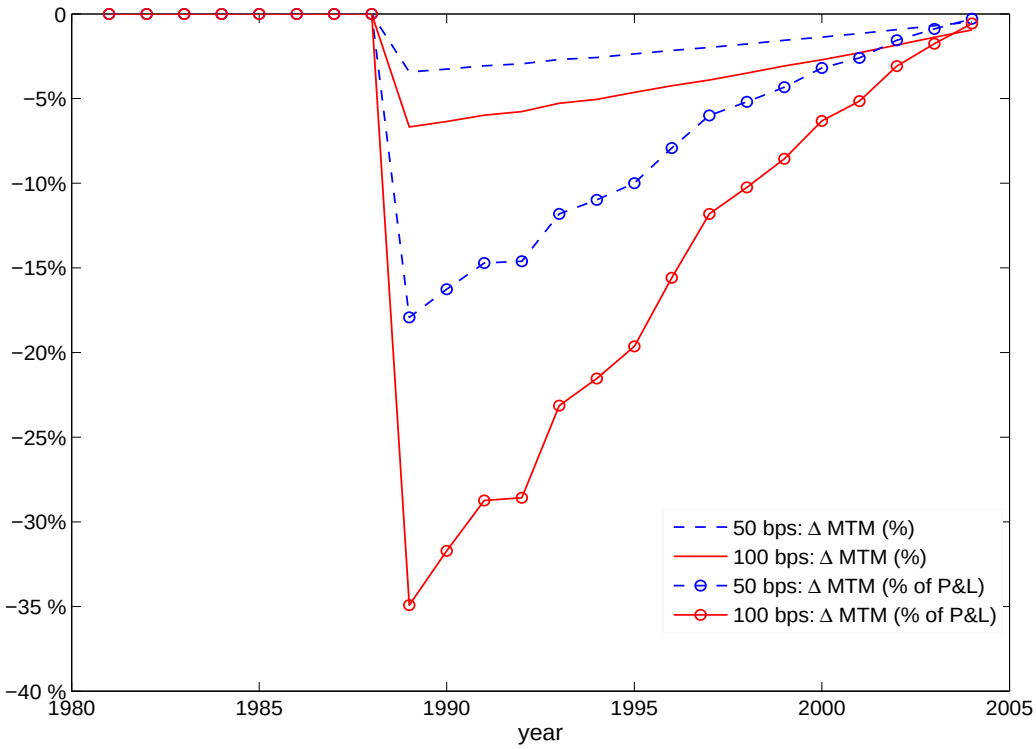


Figure 4.3: Change in mark-to-market value of the longevity swap (MTM) relative to the baseline case and the net payments from the swap, when counterparty B's credit spread widens by 50 and 100 basis points over 1988-2005.

ting default intensities[1] $(\lambda_t^{\mathrm{H}})_{t\geq 0}$, $(\lambda_t^{\mathrm{HS}})_{t\geq 0}$. Defining by $\tau := \min(\tau^{\mathrm{H}}, \tau^{\mathrm{HS}})$ the default time of the swap transaction, we further assume that, on the event $\{\tau \leq T\}$, the nondefaulting counterparty, say party i, receives a fraction $\psi^j \in [0,1]$ ($i \neq j$, with $i, j \in \{\mathrm{H}, \mathrm{HS}\}$) of the market value of the swap before default, $S_{\tau-}$, if she is in-the-money, otherwise she has to pay the full pre-default market value $S_{\tau-}$ to the defaulting counterparty. Following Duffie and Huang [1996], we can then

[1]For tractability and symmetry with the mortality model of section 2, we work with doubly stochastic default times (see the appendix). The main drawback is that the occurrence of default does not affect the conditional default probability of the surviving counterparty, thus limiting the extent to which close-out risk can be properly modelled.

write the market value of a swap with notional amount n as

$$S_0 = nE^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T (r_t + 1_{\{S_t<0\}}(1-\psi^{\mathrm{H}})\lambda_t^{\mathrm{H}} + 1_{\{S_t\geq 0\}}(1-\psi^{\mathrm{HS}})\lambda_t^{\mathrm{HS}})\mathrm{d}t\right)\left(P-\overline{p}^d\right)\right], \tag{4.3.1}$$

where P denotes the variable payment, $\overline{p}^d$ the fixed rate, and the indicator function 1_A takes the value of unity if the event A is true, zero otherwise. To understand the above formula, note that, in our setting, the risk-neutral valuation of a defaultable claim involves the use of a default-risk-adjusted short rate $r_t+\lambda_t^{\mathrm{H}}+\lambda_t^{\mathrm{HS}}$ and dividend payment $\lambda_t^{\mathrm{H}}(\psi^{\mathrm{H}}1_{S_{t-}<0}+1_{S_{t-}\geq 0})+\lambda_t^{\mathrm{HS}}(\psi^{\mathrm{HS}}1_{S_{t-}\geq 0}+1_{S_{t-}<0})$ determined by the recovery rules described above. As a result, the valuation formula (4.3.1) entails discounting at a spread above the risk-free rate given by

$$\begin{aligned}\Lambda_t :=& \lambda_t^{\mathrm{H}} + \lambda_t^{\mathrm{HS}} - \lambda_t^{\mathrm{H}}(\psi^{\mathrm{H}}1_{S_t<0} + 1_{S_t\geq 0}) - \lambda_t^{\mathrm{HS}}(\psi^{\mathrm{HS}}1_{S_t\geq 0} + 1_{S_t<0})\\ =& 1_{\{S_t<0\}}(1-\psi^{\mathrm{H}})\lambda_t^{\mathrm{H}} + 1_{\{S_t\geq 0\}}(1-\psi^{\mathrm{HS}})\lambda_t^{\mathrm{HS}},\end{aligned}$$

showing a switching-type dependence on the characteristics of the counterparty that is out-of-the-money at each given time prior to default. The swap rate admits the representation

$$\overline{p}^d = E^{\widetilde{\mathbb{P}}}[P] + \frac{\mathrm{Cov}^{\widetilde{\mathbb{P}}}\left(\exp\left(-\int_0^T (r_t+\Lambda_t)\mathrm{d}t\right), P\right)}{E^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T (r_t+\Lambda_t)\mathrm{d}t\right)\right]}, \tag{4.3.2}$$

and hence depends in a complex way not only on the interaction between the variable payments and risk factors such as interest rates, default intensities and recovery rates, but also on the path of the swap's market value itself. When P does not include a demographic component, as in the case of interest-rate swaps, the covariance term is typically negative. To see this, consider the case of the standard swap valuation formula obtained by assuming that both counterparties have the same default intensity ($\lambda_t := \lambda_t^{\mathrm{H}} = \lambda_t^{\mathrm{HS}}$) and there is no recovery conditional on default (ψ^{H} and ψ^{HS} are simply zero). If the credit risk of the counterparties is equal to the average credit quality of the LIBOR panel, the discount rate in (4.3.2) is simply given by $r+\lambda$, where λ is just the LIBOR-

Treasury (TED) spread. For a swap paying the LIBOR rate, we would then have a negative covariance term and hence $\overline{p}^d \leq E^{\widetilde{\mathbb{P}}}[P]$. When P only includes a demographic component (as in expression (4.2.3) for example), which is uncorrelated with the other variables, we would still have a non-null covariance term, due to the regime-switching nature of the discount rate in formula (4.3.2), and the fact that switching is triggered by the value of the swap, which also depends on the floating rate of interest. More generally, one might expect the covariance term to be negative, as longevity-linked payments are likely to be positively correlated with the credit quality of hedge suppliers[1] and companies with significant pension liabilities. The case of floating payments linked to both mortality and interest rates would then suggest a swap rate satisfying $\overline{p}^d \leq E^{\widetilde{\mathbb{P}}}[P]$. In the next section, we will show that this is not necessarily the case. To understand why, consider the case of full recovery as an example (set ψ^{H} and ψ^{HS} equal to one): expression (4.3.2) reduces to a default-free risk-neutral valuation formula, irrespective of both the default intensities of the counterparties and the costs involved by the credit enhancement tools needed to ensure that full recovery is indeed achieved upon default. This suggests that it is essential to consider explicitly counterparty risk mitigation tools in the pricing functional.

Counterparty risk can be mitigated in a number of ways, for example by introducing termination rights (e.g., credit puts and break clauses) or using credit derivatives (e.g., credit default swaps and credit spread options). We will focus on collateralization, a form of direct credit support requiring each party to post cash or securities when it is out-of-the-money. For simplicity, we consider the case of cash, which is by far the most common type of collateral [e.g., ISDA, 2010a] and allows us to disregard close-out risk, the risk that the value of collateral may change at default. In the interest-rate swaps market, Johannes and Sundaresan [2007] find evidence of costly collateral by comparing swap market data with swap values based on portfolios of futures and forward contracts, and by estimating a dynamic term structure model by using Treasury and swap data. We cannot carry out a similar exercise for longevity swaps, because there are no publicly available data on these transactions. On the other hand, we can quantify the

[1] This is a reasonable assumption for monoline insurers such as pension buyout firms, but might be less so for well-diversified reinsurers.

funding/opportunity costs associated with the collateral flows originating from the MTM procedure, as will be shown in section 4.5.

4.4 Collateralization

Collateral agreements reflect the amount of acceptable credit exposure that each party agrees to take on. We will consider simple collateral rules capturing the main features of the problem. Formally, let us introduce the pre-default collateral process[1] $(C_t)_{t\geq 0}$, which indicates how much cash, C_t, to post at each time t prior to default in response to changes in market conditions, including, in particular, the MTM value of the swap (we provide explicit examples below). Again, we develop our analysis from the point of view of the hedger, so that $C_t > 0$ ($C_t < 0$) means that party H is holding (posting) collateral. Using the notation $a^+ := \max(a, 0)$ and $a^- := \max(-a, 0)$, we assume the recovery rules to take the following form:

- On the event $\{\tau^{\text{H}} \leq \min(\tau^{\text{HS}}, T)\}$ (hedger's default), party HS recovers any collateral received by the hedger an instant prior to default, $C^-_{\tau^{\text{H}}-}$, and pays the full MTM value of the swap to party H if $S_{\tau^{\text{H}}-} \geq 0$. The net flow to party H is then $S^+_{\tau^{\text{H}}-} - C^-_{\tau^{\text{H}}-}$.

- On the event $\{\tau^{\text{HS}} \leq \min(\tau^{\text{H}}, T)\}$ (hedge supplier's default), party H pays the full MTM value of the swap to party HS if $S_{\tau^{\text{HS}}-} < 0$, and recovers any collateral received by HS an instant prior to default, $C^+_{\tau^{\text{HS}}-}$. The net flow to party H can then be written as $-S^-_{\tau^{\text{HS}}-} + C^+_{\tau^{\text{HS}}-}$.

- Whenever the nondefaulting counterparty, say H, is out-of-the-money, payment of the full MTM value of the swap is accomplished by party H recovering the extra amount $(S^+_{\tau-} - C^+_{\tau-})^+$ in the case of overcollateralization, or by party H paying the extra amount $(S^-_{\tau-} - C^-_{\tau-})^+$ in case of undercollateralization. In case of full collateralization, party H simply loses any collateral posted with HS.

[1] In other words, the actual collateral process supporting the transaction is $(1_{\{\tau>t\}}C_t)_{t\geq 0}$; hence, we are not concerned with the value taken by C_t after default.

To obtain neater results, it is convenient to express the collateral before default of either party as a fraction of the MTM value of the swap,

$$C_t = \left(c_t^{\text{HS}} 1_{\{S_{t-}\geq 0\}} + c_t^{\text{H}} 1_{\{S_{t-}<0\}}\right) S_{t-}, \tag{4.4.1}$$

where $c^{\text{H}}, c^{\text{HS}}$ are two nonnegative left-continuous processes giving the fraction of the MTM value of the swap that is posted as collateral by party H or HS, respectively.[1] Finally, we introduce a nonnegative continuous process $(\delta_t)_{t\geq 0}$ representing the yield on collateral, in the sense that holding/posting collateral of amount C_t yields/costs instantaneously the net amount $\delta_t C_t$ (after rebate). We can introduce some asymmetry, by setting $\delta_t = \delta_t^{\text{H}} 1_{\{S_{t-}<0\}} + \delta_t^{\text{HS}} 1_{\{S_{t-}\geq 0\}}$, so that δ_t^{H} be interpreted as party H's net cost of posting collateral when she is out-of-the-money, and δ_t^{HS} as the net yield on the collateral posted by party HS when party H is in-the-money. In general, one may regard the collateral costs embedded in swap market values as those of the marginal market participant. However, when considering individual longevity swap transactions with bespoke CSAs, it may be convenient to allow the pricing formula to take into account the cost of collateral of the counterparty: in this case δ^{H} and δ^{HS} may be regarded as the cost of posting collateral for party H and HS whenever they are out-of-the-money.

Denoting by $\overline{p}^c$ the swap rate available in case of collateralization, we can write the MTM value of the swap as in (4.3.1), but with the spread Λ now replaced by (see the appendix for a proof)

$$\Gamma_t = \lambda_t^{\text{H}}(1-c_t^{\text{H}})1_{\{S_t<0\}} + \lambda_t^{\text{HS}}(1-c_t^{\text{HS}})1_{\{S_t\geq 0\}} - \left(\delta_t^{\text{H}} c_t^{\text{H}} 1_{\{S_t<0\}} + \delta_t^{\text{HS}} c_t^{\text{HS}} 1_{\{S_t\geq 0\}}\right). \tag{4.4.2}$$

In the above expression, we recognize the typical features of valuation formulae for credit-risky securities [e.g., Bielecki and Rutkowski, 2002]: the first two terms account for the fractional recovery of the swap MTM value in case of default of the counterparty, the third one for the costs incurred when posting collateral before default. We now examine simple special cases to understand better the role of collateral in shaping swap rates.

[1] Note that representation (4.4.1) comes at a cost: we cannot encompass the case when collateral is posted by a counterparty at inception (a form of overcollateralization), which may be the case for some transactions.

4.4.1 Full collateralization

Consider the collateral rule obtained by setting c^{H} and c^{HS} equal to one, meaning that the full MTM value of the swap is received/posted as collateral depending on whether the marking-to-market process results in a positive/negative value for S_t. As we consider cash collateral, default is immaterial. In contrast with section 4.3, however, the expression for the swap MTM value does not reduce to the usual default-free, risk-neutral valuation formula in general, unless collateral costs are zero. In the case of symmetric collateral costs, for example, we obtain:

$$\overline{p}^{c} = E^{\widetilde{\mathbb{P}}}[P] + \frac{\text{Cov}^{\widetilde{\mathbb{P}}}\left(\exp\left(\int_0^T(\delta_t - r_t)\text{d}t\right), P\right)}{E^{\widetilde{\mathbb{P}}}\left[\exp\left(\int_0^T(\delta_t - r_t)\text{d}t\right)\right]}. \tag{4.4.3}$$

If the cost of collateral is positively dependent on P, we expect the swap rate to be higher than $\overline{p}^{d}$ in expression (4.3.2) [see Johannes and Sundaresan, 2007], reflecting the fact that (costly) collateralization results in the payer of the floating rate being compensated with a higher fixed rate. In the interest-rate swap market this happens for example if either the short rate or the TED spread are positively correlated with δ. The intuition is that the floating rate payer will have to both post collateral and incur higher funding costs when the floating rate increases. In longevity space, one may expect the cost of collateral to be positively dependent on mortality improvements and negatively dependent on interest rates, as longevity-linked liabilities are more capital intensive in low mortality and low interest rate environments (due to lower discounting of future cashflows). The combined impact of these two effects is ambiguous, and is discussed in the examples of section 4.5.

4.4.2 Common collateral rules

According to ISDA [2010a], it is typical for collateral agreements to specify collateral triggers based on the market value of the swap or other relevant variables (credit ratings, credit spreads, etc.) crossing pre-specified threshold levels. In longevity swaps, the CSA may define collateral rules that depend on the underlying mortality experience, involve path dependence (with respect to mortality

experience/expectations for instance), and monitor different variables at different frequency. For example, the CSA may allow for daily collateral adjustments for financial conditions, quarterly adjustments for death experience, and annual adjustments for changes in future mortality improvements. The following examples illustrate some of these aspects:

a) Set $c_t^{\text{HS}} = 1_{\{S_{t-} \geq \underline{s}(t)\}}$ and $c_t^{\text{H}} = 1_{\{S_{t-} \leq \overline{s}(t)\}}$ (for continuous functions $\underline{s}, \overline{s}$ defined on $[0, T]$ and satisfying $\underline{s} \leq \overline{s}$), meaning that the hedge supplier (hedger) is required to post full collateral if the swap's MTM value is above (below) the appropriate time-dependent threshold. More general collateral rules can be obtained by setting $c_t^{\text{HS}} = \gamma_t^{\text{HS}} 1_{\{S_{t-} \geq \underline{s}(t)\}}$ and $c_t^{\text{H}} = \gamma_t^{\text{H}} 1_{\{S_{t-} \leq \overline{s}(t)\}}$, for suitable processes $\gamma^{\text{H}}, \gamma^{\text{HS}}$ depending on prevailing market conditions or expectations about future mortality.

b) In longevity swaps, however, it is more common to define collateral thresholds in terms of mortality forecasts based on a model agreed at contract inception, and monitor the deaths in the hedger's population instead of the market value of the swap. This is due to both the re-estimation risk affecting any given mortality model and the presence of substantial model risk, which most likely would prevent the counterparties from agreeing on a common model at future dates. We can set, for example, $c_t^{\text{HS}} = 1_{\{N_{t-} \leq \alpha(t)\}}$ and $c_t^{\text{H}} = 1_{\{N_{t-} \geq \beta(t)\}}$, for continuous functions α and β satisfying $0 \leq \alpha \leq \beta \leq n$, meaning that the hedge supplier (hedger) is required to post full collateral if realized deaths are below (above) the relevant threshold.

c) For an index-based swap, it may be more convenient to work with the mortality intensity μ^I of the reference population (see section 4.2) and set $c_t^{\text{HS}} = 1_{\{\int_0^t \mu_s^I \mathrm{d}s \leq a(t)\}}$ and $c_t^{\text{H}} = 1_{\{\int_0^t \mu_s^I \mathrm{d}s \geq b(t)\}}$ for (say) continuous functions a, b satisfying $0 \leq a \leq b$. This means that collateral posting is triggered at each time t if the realized value of the longevity index, $\exp(-\int_0^t \mu_s^I \mathrm{d}s)$, falls outside the open interval $(\exp(-b(t)), \exp(-a(t)))$.

d) As was emphasized in section 4.2.1, the severity of counterparty risk depends on the credit quality of the counterparties. This is why collateralization agreements may set collateral thresholds that explicitly depend

on credit ratings or CDS spreads. A simple example of this practice can be obtained as a special case of (a) by setting $c_t^{\text{HS}} = 1_{\{N_{t-} \leq \alpha(t)\} \cup \{\lambda_t^{\text{HS}} \geq \underline{\lambda}\}}$, $c_t^{\text{H}} = 1_{\{N_{t-} \geq \beta(t)\} \cup \{\lambda_t^{\text{H}} \geq \underline{\lambda}\}}$, meaning that, at each time t, the hedger (hedge supplier) receives collateral when either realized deaths fall below the level $\alpha(t)$ (respectively $\beta(t)$) or the hedge supplier's (respectively hedger's) default intensity overshoots a given threshold $\underline{\lambda} \geq 0$. Note that both c^{H} and c^{HS} can be non zero at the same time (for example on the event $\{N_{t-} \leq \alpha(t)\} \cap \{\lambda_t^{\text{H}} \geq \underline{\lambda}\}$), but expression (4.4.1) ensures that only the party out-of-the-money will have to post collateral.

4.4.3 Computing the swap rate

The recursive nature of swap valuation formulae in the case of bilateral and asymmetric counterparty risk (in a doubly stochastic setting) was already noted by Duffie and Huang [1996]. By modeling the recovery rates and the difference in counterparties' credit spreads in reduced form, however, they could use a simple iterative procedure to determine the swap rate.[1] Here, we explicitly allow for the impact of collateral and the MTM process in the pricing functional: working in a high-dimensional Markov setting, we use a Least-Squares Monte Carlo approach. Exploiting the properties of the doubly stochastic setup, we do not model death/default times explicitly, but just rely on the mortality/default intensities [see algorithm 2 in Bacinello et al., 2010, for example]. The procedure involves the following steps (we focus on the individual forward rates for convenience):

Step 1. For an arbitrary fixed forward rate $\overline{p}_i^c \in (0,1)$, generate M simulated paths of the state variable process, X, under $\widetilde{\mathbb{P}}$ along the time grid $\mathcal{T} := \{0 < t_1, t_2, \ldots, t_n = T\}$. Denote by $S_{t_j}^{m,i}$ the MTM value of the swap and by $f_{t_j}^{i,m}$ the cashflows originating from the swap (collateral flows and swap payments) at time t_j on path m and for given forward rate $\overline{p}_i^c$.

Step 2. Compute recursively the value of the swap at time t_j (for $j = n-1, \ldots, 0$ with $t_0 = 0$) as $S_{t_j}^{m,i} = \beta_j^* \cdot e(X_{t_j}^m)$, where $e(x) := (e_1(x), \ldots, e_H(x))^\intercal$ and $\{e_1, \ldots, e_H\}$ is a finite set of functions taken from a suitable basis of $L^2(\Omega)$, and

[1] Johannes and Sundaresan [2007] sidestep recursivity issues by considering full collateralization and symmetric default risk and collateral costs.

β_j^* is given by

$$\beta_j^* = \arg\min_{\beta_j \in \mathbb{R}^H} \sum_{m=1}^{M} \left(S_{t_{j+1}}^{i,m} + f_{t_{j+1}}^{i,m} - \beta_j \cdot e(X_{t_j}^m) \right)^2 .$$

At each time t_j, use $S_{t_j}^{m,i}$ to check whether the collateral thresholds are triggered and determine the corresponding amount of collateral and associated costs.

Step 3. Iterate[1] the above procedure over different values for $\overline{p}_i^c$ until a candidate rate $\overline{p}_i^{c*}$ is found, such that the initial price, $\frac{1}{M}\sum_{m=1}^{M} S_{t_0}^{m,i*}$, is close enough to zero. Set $\overline{p}^c = \overline{p}_i^{c*}$.

Of course, the procedure relies on knowledge of the dynamics of the state variable process under the pricing measure. To this end, in the next section, we outline a calibration approach based on the joint use of fixed-income data and funding costs / capital requirements for longevity-linked liabilities.

4.5 Examples

We use a continuous-time model for the risk-free yield curve, the LIBOR and mortality rates, as well as for the cost of collateral. The credit risk of party HS is assumed to be equal to the average credit quality of the LIBOR panel, so that the TED spread would be party HS's default intensity if there were zero recovery upon default (see section 4.3). We then set $\lambda^{\text{H}} = \lambda^{\text{HS}} + \Delta$ and consider two cases: party H is either of the same credit quality as party HS ($\Delta = 0$) or is more credit-risky ($\Delta > 0$).

We describe the evolution of uncertainty by a six-dimensional state variable vector X with the Gaussian dynamics reported in appendix C.2. The first four components are: the short rate, $r = X^{(1)}$, assumed to revert to the long-run central tendency factor $X^{(2)}$, representing the slope of the risk-free yield curve; the TED spread $X^{(3)}$, so that the LIBOR rate is given by $X^{(1)} + X^{(3)}$; and the net yield on collateral in the interest-rate swap market, $X^{(4)}$. The latter factor is used to draw a comparison with the cost of collateral in the longevity swap market.

[1] In the numerical examples of section 4.5, we use a combination of bisection, secant, and inverse quadratic interpolation methods to compute $\overline{p}_i^{c*}$ [see Forsythe et al., 1976].

The remaining two components describe the opportunity cost of longevity swap dealers, $X^{(5)}$, and the log-intensity of mortality of a given population, $\log \mu = X^{(6)}$. Under the assumption of independence between $(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)})$ and $X^{(6)}$, we can estimate separately the dynamics of the two groups of factors. For the first vector, we rely on the estimates of Johannes and Sundaresan [2007], who use weekly Treasury and swap data from 1990 to 2002 to obtain the parameter values reported in table 4.2. For the intensity $\exp(X_t^{(6)})$, we use a continuous-time version of the Lee-Carter mortality projection model for a cohort of 65-year olds; see appendix B for details.

As a first example, we focus on funding costs and simply take $\delta^{\text{HS}} = X^{(3)}$ and $\delta^{\text{H}} = X^{(3)} + \Delta$, meaning that the hedger's net cost of collateral coincides with its funding costs net of the short rate (assuming it is rebated), whereas the hedger's net yield on the collateral amounts posted by party HS coincides with the TED spread. Assuming that the pricing formula uses information on the collateral costs of the counterparty, an alternative interpretation is that each party's net collateral costs coincide with their borrowing costs net of the risk-free rate. In the case of asymmetric default risk, we consider values of 100 and 200 basis points for Δ.

We compute the longevity swap rates for a 25-year swap written on a population of 10,000 US males aged 65 at the beginning of 2008. In figure 4.4, we plot the underlying forward rates obtained for different collateralization rules against the percentiles of survival rate improvements based on Lee-Carter forecasts. We see that margins are positive and increasing with payment maturity in the case of symmetric default risk, for both uncollateralized and fully collateralized transactions. As soon as we introduce asymmetry in default risk ($\Delta > 0$), however, margins widen in the case of no collateralization, reflecting the fact that the hedger needs to pay an additional premium on account of its higher credit risk. In the case of full collateralization, counterparty risk is neutralized, but the hedger is compensated for her higher funding costs and the positive dependence between funding costs and collateral amounts discussed before: equilibrium swap rates are pushed lower and produce a negative margin on best estimate swap rates.

In figure 4.5, we examine the swap margins induced by one-way collateral-

κ_1	0.969	η_1	-0.053	σ_1	0.008	UK	
κ_2	0.832	η_3	-0.014	σ_2	0.155	δ_K	-0.888
κ_3	1.669	η_4	0.007	σ_3	0.009	σ_K	1.156
κ_4	0.045	η_5	0.055	σ_4	0.010	US	
κ_5	0.990	$\kappa_{5,1}$	0.147	σ_5	0.690	δ_K	-0.761
$\kappa_{3,1}$	-0.163	$\kappa_{5,2}$	1.340	θ_2	0.046	σ_K	1.078
$\kappa_{4,1}$	0.114	$\kappa_{5,3}$	2.509	θ_3	0.003		
$\kappa_{3,4}$	0.804	$\kappa_{5,4}$	-0.133	θ_4	0.007		
$\kappa_{4,2}$	-0.038	$\kappa_{5,6}$	-0.002	θ_5	0.115	$\rho_{1,2}$	-0.036

Table 4.2: Parameter values for the dynamics of X given in Appendix C.2. The estimates for $X^{(5)}$ are based on the assumption that capital increases are funded by counterparties at 6% plus the LIBOR rate.

ization in the case of asymmetric default risk. When only the hedge supplier has to post full collateral, forward rates are higher than best estimate survival probabilities, meaning that the hedger has to compensate the hedge supplier for bearing both the cost of risk mitigation and the hedger's higher default risk. The opposite is true when it is the hedger who has to post full collateral when out-of-the money. In this case, swap margins are clearly negative, and decreasing in payment maturity. These effects are amplified when the asymmetry in counterparties' credit quality is greater, as can be seen from the spreads reported in table 4.3 for some key maturities and collateralization rules.

Plotting the swap rate *margins* against best estimate mortality improvements allows one to interpret the swap rates as outputs of a pricing functional based on adjustments to a reference mortality model (which is common practice in longevity space). On the other hand, longevity swap *spreads* are easier to compare with those emerging in other transactions. In table 4.4, we make a comparison with the interest-rate swap spreads implied by our parameterization of the state vector $(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)})$. In particular, we report the difference between interest-rate futures prices (obtained by considering full collateralization and setting the cost of collateral equal to the risk-free rate) and interest-rate forward rates for collateralized transactions with collateral costs equal to the funding costs of the counterparties. Spreads are negative, in line with the intuition that interest rate risk leads to a discount for the payer of the fixed rate, as discussed in the introduction, and are of a magnitude consistent with the findings of Jo-

hannes and Sundaresan [2007]. The results show that longevity swap spreads are comparable with, and often much smaller (in absolute value) than, those found in the interest-rate swap market. For example, in the case of bilateral full collateralization, longevity forward rates for 15- to 25-year maturities embed a spread substantially smaller than that of interest-rate forwards of corresponding maturity. In the case of one-way collateralization on the hedger's side, in interest-rate forward rates we find a discount (negative spread) that turns into a premium (positive spread) of comparable size in the corresponding longevity swap, due to the additional and opposite effect of longevity risk on swap rates. Our findings are robust to the choice of maturity, collateralization rules, and counterparty credit quality, and are mainly driven by two effects: i) the different nature of the risk underlying the swap, a survival curve in the case of longevity swaps, and a floating rate in the case of interest-rate swaps; ii) the fact that interest rate risk and longevity risk impact longevity swap margins in opposite directions, thus diluting the overall effect of collateralization on longevity swap rates.

$\lambda^{\text{H}} = \lambda^{\text{HS}} + \Delta$ $\delta^{\text{H}} = \delta^{\text{HS}} + \Delta$	Maturity payment (yrs)	$c^{\text{H}} = 0$ $c^{\text{HS}} = 0$ (bps)	$c^{\text{H}} = 0$ $c^{\text{HS}} = 1$ (bps)	$c^{\text{H}} = 1$ $c^{\text{HS}} = 0$ (bps)	$c^{\text{H}} = 1$ $c^{\text{HS}} = 1$ (bps)
$\Delta = 0$	15	0.03	11.34	-11.76	0.05
	20	1.11	19.93	-17.94	0.86
	25	1.50	21.25	-18.35	1.24
$\Delta = 100$ bps	15	5.45	16.79	-17.29	-5.84
	20	10.16	28.95	-27.08	-8.23
	25	10.96	30.75	-27.76	-9.19
$\Delta = 200$ bps	15	11.30	22.29	-22.90	-11.25
	20	19.26	38.06	-36.16	-17.42
	25	19.46	40.27	-37.02	-18.38

Table 4.3: Forward rate spreads $\overline{p}^{c}_{T_i} - p_{T_i}$ (in basis points) for different collateralization rules, maturities and spread $\Delta \in \{0, 0.01, 0.02\}$. The LSMC procedure uses 5000 paths over a quarterly grid with polynomial basis functions of order 3, and is repeated for 100 seeds.

In a second example, we focus on the opportunity cost of selling additional longevity protection. As we do not have any publicly available transaction data from the longevity swap market to calibrate $X^{(5)}$, we simulate the capital charges arising from holding a representative longevity-linked liability in response to changes in the evolution of the factors $(X^{(1)}, \ldots, X^{(4)})$ and $X^{(6)}$. We therefore

$\lambda^{\text{H}} = \lambda^{\text{HS}} + \Delta$ $\delta^{\text{H}} = \delta^{\text{HS}} + \Delta$	Maturity payment (yrs)	IRS ($\delta^{\text{HS}} = X^{(4)}$) $c^{\text{H}} = 0$ $c^{\text{HS}} = 1$ (bps)	$c^{\text{H}} = 1$ $c^{\text{HS}} = 0$ (bps)	$c^{\text{H}} = 1$ $c^{\text{HS}} = 1$ (bps)	Longevity ($\delta^{\text{HS}} = X^{(3)}$) $c^{\text{H}} = 0$ $c^{\text{HS}} = 1$ (bps)	$c^{\text{H}} = 1$ $c^{\text{HS}} = 0$ (bps)	$c^{\text{H}} = 1$ $c^{\text{HS}} = 1$ (bps)
$\Delta = 0$	15	-7.96	-44.97	-52.86	11.34	-11.76	0.05
	20	-12.68	-42.64	-56.22	19.93	-17.94	0.86
	25	-17.94	-40.98	-58.92	21.25	-18.35	1.24
$\Delta = 100$ bps	15	-8.00	-67.87	-75.23	16.79	-17.29	-5.84
	20	-12.65	-63.84	-77.42	28.95	-27.08	-8.23
	25	-17.65	-60.63	-77.64	30.75	-27.76	-9.19

Table 4.4: Comparison of interest-rate swaps (IRSs) with longevity swaps. The IRS spreads represent the difference betweeen the futures prices (the opportunity cost of collateral coincides with the risk-free rate for both parties) and the forward rate for a collateralized transaction (for different collateralization rules, maturities, and credit risk).

'synthesize' the realizations of $X^{(5)}$ by using information on regulatory requirements to quantify the capital charges accruing to the counterparties during the life of the swap. In particular, we use the following bottom-up procedure:

Step 1: We simulate several paths of the factors $X^{(1)}, \ldots, X^{(4)}$ and $X^{(6)}$ along a time grid $\hat{\mathfrak{T}} := \{t_1, t_2, \ldots, t_k\}$ (with $t_k = \hat{T} > t_1 > 0$) and under the pricing measure $\widetilde{\mathbb{P}}$. Again, for our example, we assume the $\widetilde{\mathbb{P}}$-dynamics of $X^{(6)}$ to be the same as under the physical measure.

Step 2: The paths simulated in the previous step are used to compute, at each date $t \in \hat{\mathfrak{T}}$, the regulatory capital needed by an insurer to hold the liability $n - N_{t+\overline{T}}$, where $\overline{T} < \hat{T}$ is a representative maturity proxying the average duration of longevity-linked liabilities in the longevity swap market. We use $\overline{T} = 15$ and $\hat{T} = 40$ (years) for our example. To compute the capital requirements, we use the Solvency II framework, which is based on the 99.5% value-at-risk of the net assets over a one-year horizon. For simplicity, we assume holders of longevity exposures to be invested in cash. The distribution of the one-year-ahead market-consistent value of the liability usually requires nested simulation, unless a simplified approach is adopted. In our setting, market-consistent discount factors can be computed analytically based on the one-year-ahead simulated realizations, as the pair $(X^{(1)}, X^{(2)})$ is an affine process. We use the LSMC approach (see

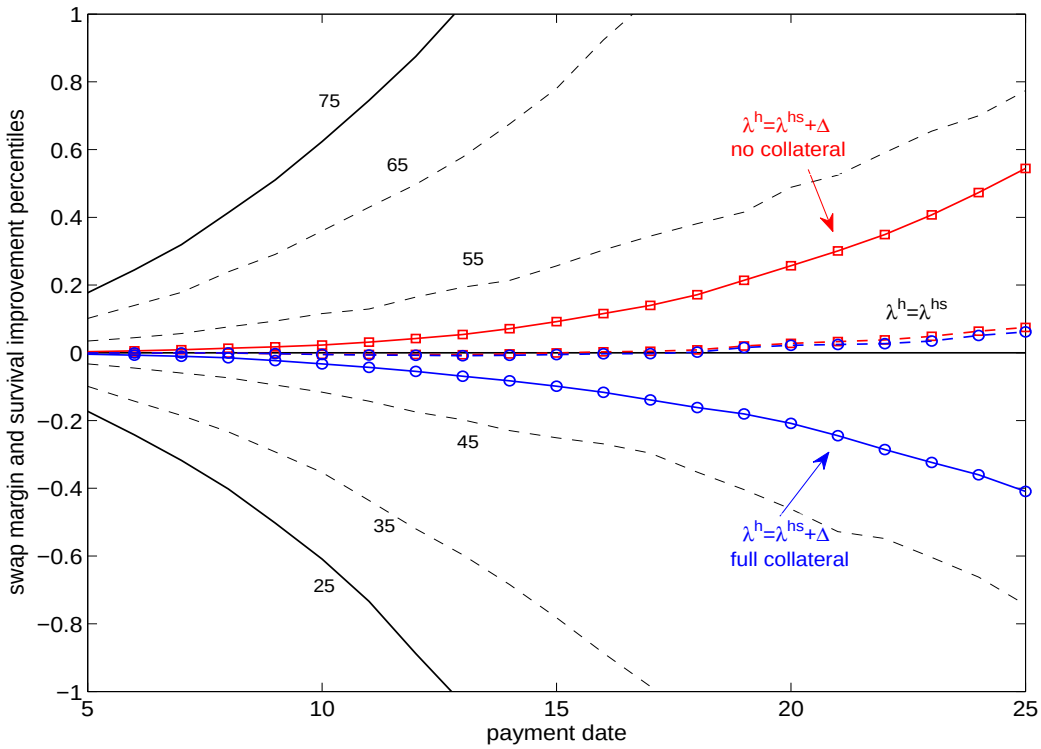


Figure 4.4: Swap margins $\overline{p}^c_{T_i}/p_{T_i} - 1$ computed for different maturities $\{T_i\}$ and collateral rules, with $\delta^{\mathrm{H}} = \lambda^{\mathrm{H}}$, $\delta^{\mathrm{HS}} = \delta^{\mathrm{H}} + \Delta$, and $\lambda^{\mathrm{H}} = \lambda^{\mathrm{HS}} + \Delta$, with $\Delta = 0$ (dashed lines) or $\Delta = 0.01$ (solid lines): no collateral (squares), full collateralization (circles). The underlying is a cohort of 10,000 US males aged 65 at the beginning of 2008. Forward rates are plotted against the percentiles of improvements in survival rates based on Lee-Carter forecasts.

section 4.4.3) to determine the expected number of survivors.[1]

Step 3: We use the simulated capital charges obtained in the previous step to compute the gains/costs incurred to reduce/increase capital at each time step along each simulated path. We assume that capital charges are funded at the counterparties' funding cost, plus a spread of 6%[2] to reflect the opportunity cost of diverting to an individual liability funds that could be used to support insurance

[1] See Stevens et al. [2010a] for other approximation methods in the context of Lee-Carter forecasts.

[2] This is a reasonable, conservative value for the return on capital of longevity swaps dealers: anecdotal evidence suggests that it can be twice as large.

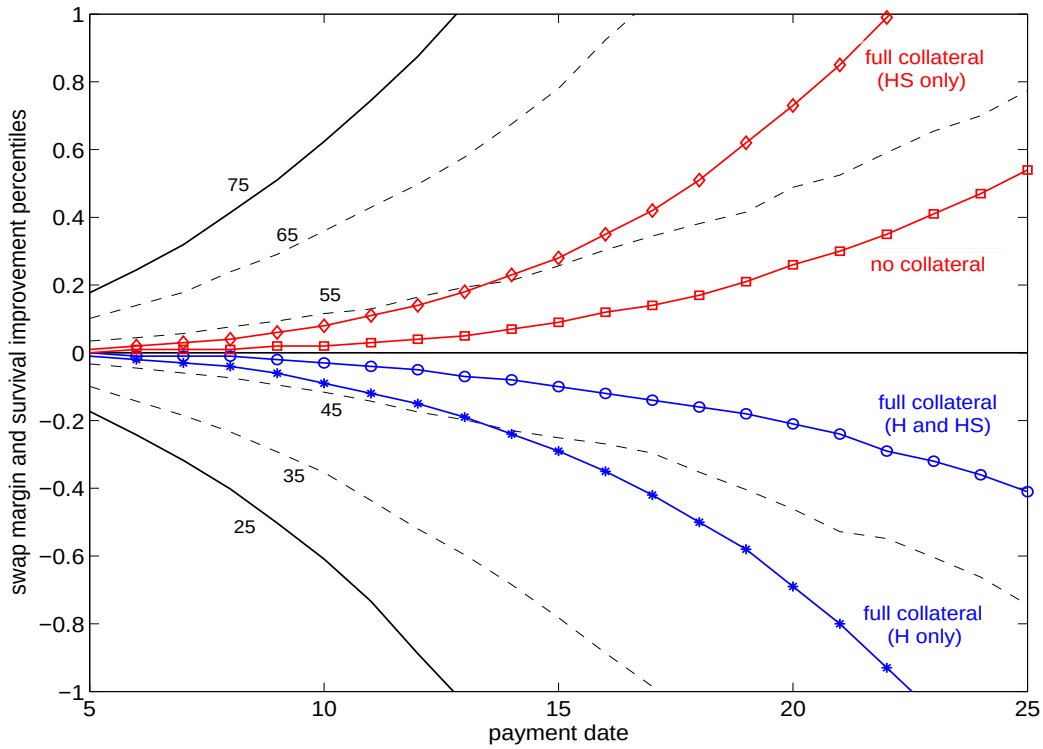


Figure 4.5: Swap margins $\overline{p}^c_{T_i}/p_{T_i} - 1$ computed for different maturities $\{T_i\}$ and collateral rules, with $\delta^{\mathrm{H}} = \lambda^{\mathrm{H}}$, $\delta^{\mathrm{HS}} = \delta^{\mathrm{H}} + 0.01$, and $\lambda^{\mathrm{H}} = \lambda^{\mathrm{HS}} + 0.01$: no collateral (squares), full collateralization (circles), full collateral posted only by party H (stars) or party HS (diamonds). The underlying is a cohort of 10,000 US males aged 65 at the beginning of 2008. Forward rates are plotted against the percentiles of improvements in survival rates based on Lee-Carter forecasts.

business at the aggregate level. The simulated realizations of the opportunity cost of capital (see figure 4.7 for an example) are used to estimate the dynamics of $X^{(5)}$ reported in the appendix. The parameter estimates are included in table 4.2.

In the case of symmetric collateralization, we find results comparable with those obtained by using the counterparties' funding costs for the process δ. However, figure 4.6 shows that margins increase (decrease) considerably when one-way collateralization on the hedge supplier's (hedger's) side is considered. This is because the party required to post collateral explicitly takes into account tail events in computing collateral costs, whereas in figure 4.5 funding costs where computed

on the basis of the market value of the longevity swap.

Finally, we study the sensitivity of longevity forward spreads to the volatility of the net collateral cost $X^{(5)}$. To close off the interest-rate risk channel, we fix the factors $X^{(1)}, X^{(2)}$ equal to their long-run means. Table 4.5 reports the results obtained for different values of the volatility parameter σ_5 in the case of symmetric default risk and bilateral full collateralization. We see that spreads increase dramatically for large values of the volatility parameter, but are comparable with those found in the previous examples for reasonable volatility levels (i.e., below 5%).

σ_5	p_{25}	$\overline{p}^c$	spread (bps)
0.0005	0.201425	0.201469	2.15
0.0100	0.201425	0.201822	19.68
0.0150	0.201425	0.202009	28.96
0.0200	0.201425	0.202196	38.26
0.1000	0.201425	0.205237	189.24
0.1500	0.201425	0.207184	285.90

Table 4.5: Sensitivity with respect to parameter σ_5: we compute 25-year forward rates and spreads (in basis points) under full collateralization by setting $X^{(1)}, X^{(2)}$ equal to their long run means. The baseline estimated parameter values for the dynamics of $X^{(5)}$ are $\theta_5 = 0.000254$, $\kappa_5 = 1.005073$, $\sigma_5 = 0.000542$, $\eta_5 = 0.000269$, $\kappa_{53} = 0.003648$, $\kappa_{54} = 0.000018$, $\kappa_{56} = 0.000261$.

4.6 Conclusion

In this study, we have provided a framework for understanding and quantifying the cost of bilateral default risk and collateral strategies on longevity risk solutions. The results address the concerns aired by potential hedgers regarding how to measure the trade-off between the hedge effectiveness of longevity-linked instruments and the counterparty risk they involve. We have described a methodology for pricing longevity swaps that explicitly takes into account the dynamics of the marking-to-market process, the collateral flows it generates, and the costs associated with the posting of collateral. We have shown how collateral strategies can mitigate if not eliminate counterparty risk, but inevitably introduce an extra

cost that must be borne by the hedge supplier or by the hedger, depending on how their credit quality and collateral costs compare with each other. Our most significant and useful finding is that the overall cost of the collateralization strategies in the longevity swap market is comparable with, and often smaller than, that found in the much more liquid interest-rate swap market. Hence, there is no reason to suppose that counterparty risk will provide an insurmountable barrier to the further development of the longevity swap market. Our analysis accordingly provides a robust framework for comparing the costs of credit enhancement in bespoke longevity swaps with the benefits offered by competing solutions such as securitization and indexed swaps.

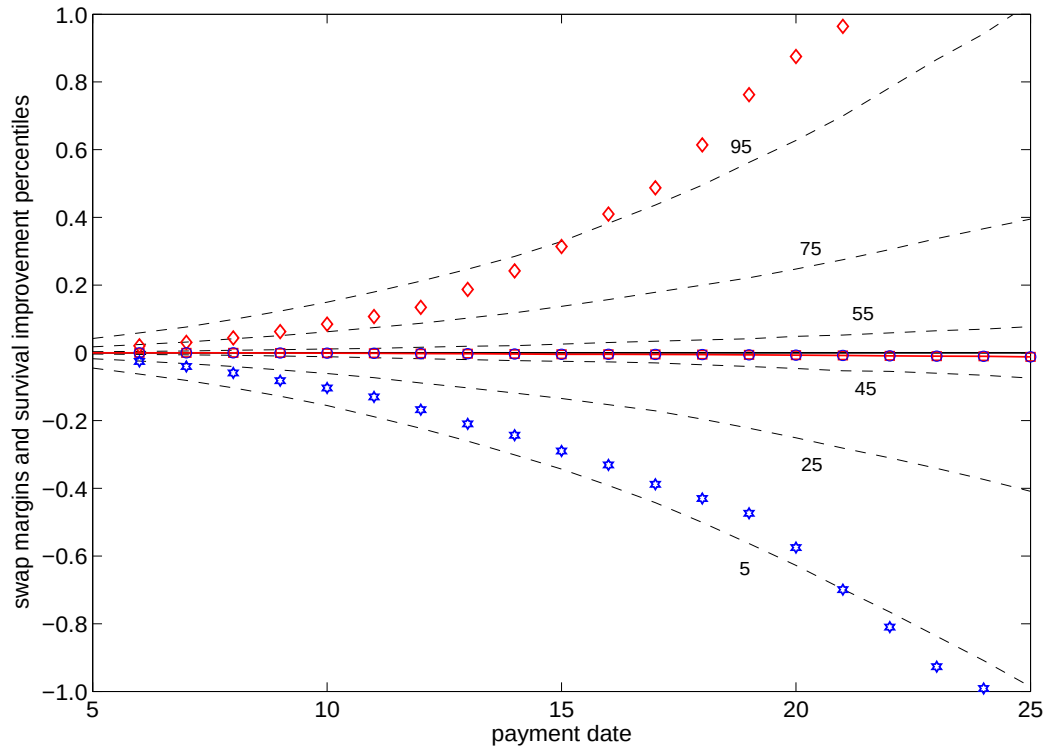


Figure 4.6: Swap margins $\overline{p}^c_{T_i}/p_{T_i} - 1$ computed for different maturities $\{T_i\}$ and collateral rules, with $\lambda^{\text{H}} = \lambda^{\text{HS}}$ and $\delta^{\text{H}} = \delta^{\text{HS}} = X^{(5)}$, where the parameter estimates for the dynamics of $X^{(5)}$ are given in table 4.2. Collateral rules: no collateral (squares), full collateralization (circles), full collateral posted only by party H (stars) or HS (diamonds). Forward rates are plotted against the percentiles of improvements in survival rates based on Lee-Carter forecasts (65-year old US males in 2008).

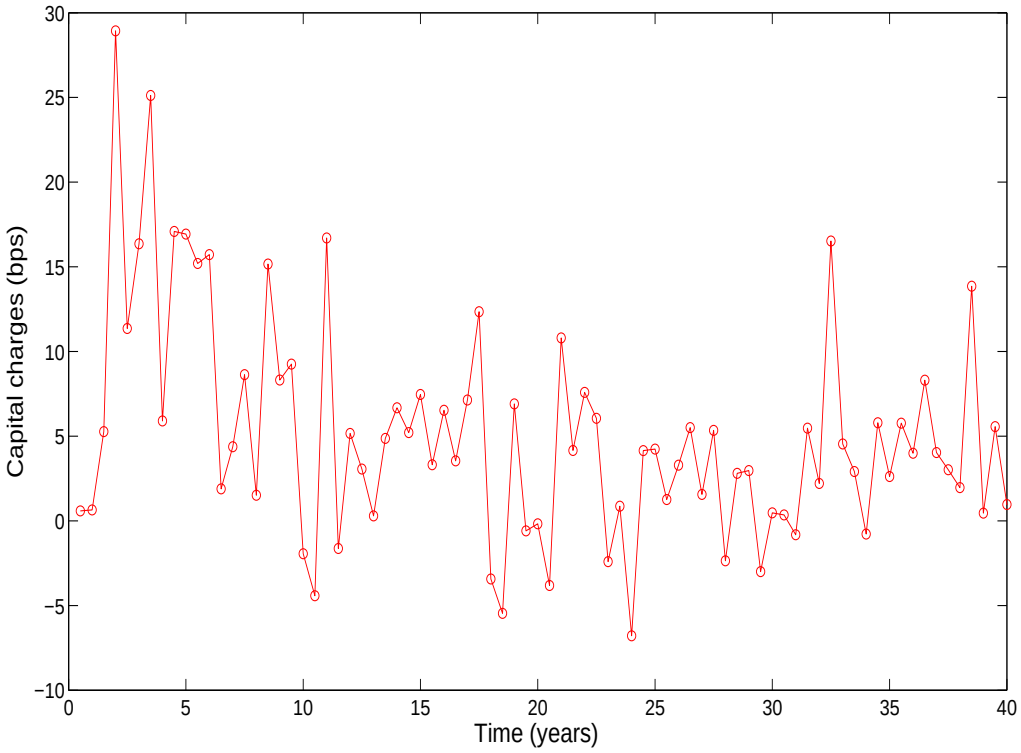


Figure 4.7: A simulated path of the capital charges accruing to the longevity swap dealer holding a representative longevity-linked liability $n - N_{t+\overline{T}}$ under the Solvency II regulatory framework.

Part IV

Appendices

Appendix A

Chapter 2

A.1 Proof of (2.1.3)

Case of Realized Covariance (RC) estimator:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\lim_{h\to 0^+} \mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1+h}\right)\right] &= \mathbb{V}\mathrm{ar}\left[\mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1}\right)\right] \\
&= \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N}\left(X^{(1)}_{t^{(1)}_i} - X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(1)}_i} - X^{(2)}_{t^{(1)}_{i-1}}\right)\right] \\
&= \sum_{i=1}^{N}\mathbb{V}\mathrm{ar}\left[\left(X^{(1)}_{t^{(1)}_i} - X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(1)}_i} - X^{(2)}_{t^{(1)}_{i-1}}\right)\right] \\
&= \sum_{i=1}^{N}\left(t^{(1)}_i - t^{(1)}_{i-1}\right)^2\left(\sigma^2_{1,2} + \sigma_{1,1}\,\sigma_{2,2}\right) \\
&= \left(\sigma^2_{1,2} + \sigma_{1,1}\,\sigma_{2,2}\right)\sum_{i=1}^{N}\left(t^{(1)}_i - t^{(1)}_{i-1}\right)^2 \quad ;
\end{aligned}$$

Assuming that, for $i = 1, \ldots, N$, $t^{(1)}_i - t^{(1)}_{i-1} = \Delta$, then we can have

$$\mathbb{V}\mathrm{ar}\left[\lim_{h\to 0^+} \mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1}, X^{(2)}_{\Pi_1+h}\right)\right] = \Delta^2\, N\, \left(\sigma^2_{1,2} + \sigma_{1,1}\,\sigma_{2,2}\right) \ .$$

Case of Hayashi-Yoshida (HY) estimator:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\lim_{h\to 0^+}\mathrm{HY}\left(X^{(1)}_{\Pi_1},X^{(2)}_{\Pi_1+h}\right)\right] =& \mathbb{V}\mathrm{ar}\left[\lim_{h\to 0^+}\sum_{i,j}\left(X^{(1)}_{t^{(1)}_i}-X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(1)}_j+h}-X^{(2)}_{t^{(1)}_{j-1}+h}\right)\right.\\
&\left.1_{\left\{\left(t^{(1)}_i,t^{(1)}_{i-1}\right]\cap\left(t^{(1)}_j+h,t^{(1)}_{j-1}+h\right]\neq\varnothing\right\}}\right]\\
=& \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N}\left(X^{(1)}_{t^{(1)}_i}-X^{(1)}_{t^{(1)}_{i-1}}\right)\left\{\left(X^{(2)}_{t^{(1)}_i}-X^{(2)}_{t^{(1)}_{i-1}}\right)\right.\right.\\
&\left.\left.+\left(X^{(2)}_{t^{(1)}_{i+1}}-X^{(2)}_{t^{(1)}_i}\right)\right\}\right]\\
=& \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N}\left(X^{(1)}_{t^{(1)}_i}-X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(1)}_i}-X^{(2)}_{t^{(1)}_{i-1}}\right)\right]\\
&+\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N-1}\left(X^{(1)}_{t^{(1)}_i}-X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(1)}_{i+1}}-X^{(2)}_{t^{(1)}_i}\right)\right]\\
=& \mathbb{V}\mathrm{ar}\left[\mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1},X^{(2)}_{\Pi_1}\right)\right]\\
&+\sum_{i=1}^{N-1}\mathbb{V}\mathrm{ar}\left[\left(X^{(1)}_{t^{(1)}_i}-X^{(1)}_{t^{(1)}_{i-1}}\right)\left(X^{(2)}_{t^{(1)}_{i+1}}-X^{(2)}_{t^{(1)}_i}\right)\right]\\
=& \left(\sigma^2_{1,2}+\sigma_{1,1}\,\sigma_{2,2}\right)\sum_{i=1}^{N}\left(t^{(1)}_i-t^{(1)}_{i-1}\right)^2+\sigma_{1,1}\,\sigma_{2,2}\sum_{i=1}^{N-1}\left(t^{(1)}_i-t^{(1)}_{i-1}\right)^2
\end{aligned}$$

Again, assuming that, for $i=1,\ldots,N$, $t^{(1)}_i-t^{(1)}_{i-1}=\Delta$, it becomes

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\lim_{h\to 0^+}\mathrm{RC}_{\Pi_1}\left(X^{(1)}_{\Pi_1},X^{(2)}_{\Pi_1+h}\right)\right] =& \Delta^2\,N\,\left(\sigma^2_{1,2}+\sigma_{1,1}\,\sigma_{2,2}\right)+\Delta^2\,(N-1)\,\sigma_{1,1}\,\sigma_{2,2}\\
=& \Delta^2\left[N\,\sigma^2_{1,2}+(2\,N-1)\,\sigma_{1,1}\,\sigma_{2,2}\right]
\end{aligned}$$

Appendix B

Chapter 3

B.1 Proof of LW Bias Correction when the Drift is Estimated in Sample

B.1.1 Deterministic drift process

See section B.1.2.

B.1.2 Generalization to OU-type Processes

The proof in this section can be applied to the case of a deterministic drift process by setting $\kappa_l = 0$ and $a_l = 0$, $l = 1, 2$.

Additional Notation Used In the Proofs

$\Delta X_{1,i}^{(l)}$ Observed increment of process $X^{(l)}$ over the interval $I_i^{(l)}$, as calculated by the calibration equation of the model (BM and OU types).

$$\begin{aligned}\Delta Y_i^{(l)} &:= X_{t_i^{(l)}}^{(l)} - e^{-\kappa_l\left(t_i^{(l)} - t_{i-1}^{(l)}\right)} X_{t_{i-1}^{(l)}}^{(l)} - a_l\left(1 - e^{-\kappa_l\left(t_i^{(l)} - t_{i-1}^{(l)}\right)}\right) \\ &= X_{t_i^{(l)}}^{(l)} - e^{-\kappa_l\left(t_i^{(l)} - t_{i-1}^{(l)}\right)} X_{t_{i-1}^{(l)}}^{(l)} - \left|I_i^{(l)}\right|_a \kappa_l a_l \\ &= \left|I_i^{(l)}\right|_a \mu_l + \sigma_l \int_{t_{i-1}^{(l)}}^{t_i^{(l)}} e^{-\kappa_l\left(t_i^{(l)} - t\right)} \mathrm{d}W_t^{(1)} \ ,\end{aligned}$$

Then the distribution of $\Delta Y_i^{(l)}$ is given by

$$\begin{aligned}\Delta Y_i^{(l)} &\sim N\left(\left|I_i^{(l)}\right|_a \mu_l,\ \sigma_l^2 \frac{1}{2\kappa_l}\left(1-e^{2\kappa(t_2-t_1)}\right)\right) \\ &\sim N\left(\left|I_i^{(l)}\right|_a \mu_l,\ \left|I_i^{(l)}\right|_d \sigma_l^2\right)\ .\end{aligned}$$

Preliminary Results

Mean Estimator (MLE)

Unbiasedness of $\overline{\mu}_l$. Let $\gamma_i := \gamma_{t_i^{(1)}}$ and $\delta_j := \delta_{t_j^{(2)}}$. The unbiasedness of the mean estimator can be checked as follows

$$\begin{aligned}\mathbb{E}\left[\overline{\mu}_1\right] &= \mathbb{E}\left[\frac{1}{\psi}\sum_i \gamma_i \Delta Y_i^{(1)}\right] \\ &= \frac{1}{\psi}\sum_i \gamma_i \left|I_i^{(1)}\right|_a \mu_1 \\ &= \mu_1\ .\end{aligned}$$

The unbiasedness of $\overline{\mu}_2$ can be checked similarly. □

Assuming Zero Mean When we assume increments have mean of zero, the LW estimator is given by

$$\Gamma\left(\xi^{(1)},\xi^{(2)}\right) := \frac{1}{\phi}\sum_{i,j}\Delta Y_i^{(1)}\Delta Y_j^{(2)} w_{ij}\ .$$

Unbiasedness of $\Gamma\left(\xi^{(1)},\xi^{(2)}\right)$. In order to check the unbiasedness of the estimator

we verify that $\mathbb{E}\left[\Gamma\left(\xi^{(1)},\xi^{(2)}\right)\right]=\sigma_{1,2}$:

$$\begin{aligned}
\mathbb{E}\left[\Gamma\left(\Delta Y^{(1)},\Delta Y^{(2)}\right)\right] &= \mathbb{E}\left[\frac{1}{\phi}\sum_{i,j}\Delta Y_i^{(1)}\Delta Y_j^{(2)}w_{ij}\right] \\
&= \frac{1}{\phi}\sum_{i,j}\mathbb{E}\left[\Delta Y_i^{(1)}\Delta Y_j^{(2)}\right]w_{ij} \\
&= \frac{1}{\phi}\sum_{i,j}\left|I_i^{(1)}\cap I_j^{(2)}\right|_c\sigma_{1,2}\,w_{ij} \\
&= \sigma_{1,2}\frac{1}{\phi}\sum_{i,j}\frac{\left|I_i^{(1)}\cap I_j^{(2)}\right|_d}{\left|I_i^{(1)}\right|_d}\frac{\left|I_i^{(1)}\cap I_j^{(2)}\right|_o}{\left|I_j^{(2)}\right|_o} \\
&= \sigma_{1,2}\ .
\end{aligned}$$

□

NOT Assuming Zero Mean When we do not assume that increments have mean of zero, the LW estimator is given by

$$\Gamma\left(\xi^{(1)},\xi^{(2)}\right):=\ \frac{1}{\phi-\eta}\ \sum_{i,j}\left(\Delta Y_i^{(1)}-\left|I_i^{(1)}\right|_a\overline{\mu}_1\right)\left(\Delta Y_j^{(2)}-\left|I_j^{(2)}\right|_b\overline{\mu}_2\right)w_{ij}\ .$$

Unbiasedness of $\Gamma\left(\xi^{(1)},\xi^{(2)}\right)$. In order to check the unbiasedness of the estimator

we need to verify that $\mathbb{E}\left[\Gamma\left(\xi^{(1)},\xi^{(2)}\right)\right]=\sigma_{1,2}$. Let $\gamma_i:=\gamma_{t_i^{(1)}}$ and $\delta_j:=\delta_{t_j^{(2)}}$, then

$$\begin{aligned}
\mathbb{E}\left[\Gamma\left(\xi^{(1)},\xi^{(2)}\right)\right] =&\mathbb{E}\left[\frac{1}{\phi-\eta}\sum_{i,j}\left(\Delta Y_i^{(1)}-\left|I_i^{(1)}\right|_a\overline{\mu}_1\right)\left(\Delta Y_j^{(2)}-\left|I_j^{(2)}\right|_b\overline{\mu}_2\right)w_{ij}\right]\\
=&\frac{1}{\phi-\eta}\mathbb{E}\Bigg[\sum_{i,j}w_{ij}\left(\Delta Y_i^{(1)}-\frac{\left|I_i^{(1)}\right|_a}{\psi}\sum_h\gamma_h\Delta Y_h^{(1)}\right)\\
&\left(\Delta Y_j^{(2)}-\frac{\left|I_j^{(2)}\right|_b}{\varphi}\sum_k\delta_k\Delta Y_k^{(2)}\right)\Bigg]\\
=&\frac{1}{\phi-\eta}\mathbb{E}\Bigg[\sum_{i,j}w_{ij}\Bigg\{\Delta Y_i^{(1)}\ \Delta Y_j^{(2)}-\ \alpha_i\ \Delta Y_j^{(2)}\sum_h\gamma_h\Delta Y_h^{(1)}\\
&-\ \beta_j\ \Delta Y_i^{(1)}\sum_k\delta_k\Delta Y_k^{(2)}+\ \alpha_i\ \beta_j\sum_h\gamma_h\Delta Y_h^{(1)}\sum_k\delta_k\Delta Y_k^{(2)}\Bigg\}\Bigg]\ ,
\end{aligned}$$

where $\alpha_i=\frac{\left|I_i^{(1)}\right|_a}{\psi}$ and $\beta_j=\frac{\left|I_j^{(2)}\right|_b}{\varphi}$,

$$\begin{aligned}
=&\frac{1}{\phi-\eta}\sum_{i,j}w_{ij}\,\mathbb{E}\left[\Delta Y_i^{(1)}\ \Delta Y_j^{(2)}\right]-\sum_{i,j}w_{ij}\,\alpha_i\,\mathbb{E}\left[\Delta Y_j^{(2)}\sum_h\gamma_h\Delta Y_h^{(1)}\right]\\
&-\ \sum_{i,j}w_{ij}\,\beta_j\,\mathbb{E}\left[\Delta Y_i^{(1)}\sum_k\delta_k\Delta Y_k^{(2)}\right]+\ \sum_{i,j}w_{ij}\,\alpha_i\,\beta_j\,\mathbb{E}\left[\sum_h\gamma_h\Delta Y_h^{(1)}\sum_k\delta_k\Delta Y_k^{(2)}\right]\quad ,\\
=&\frac{1}{\phi-\eta}\left\{\sum_{i,j}w_{ij}\ \mathbf{A}_{ij}^{(\mathbf{1})}-\sum_{i,j}w_{ij}\ \alpha_i\ \mathbf{A}_{ij}^{(\mathbf{2})}-\sum_{i,j}w_{ij}\ \beta_j\ \mathbf{A}_{ij}^{(\mathbf{3})}+\sum_{i,j}w_{ij}\ \alpha_i\ \beta_j\ \mathbf{A}_{ij}^{(\mathbf{4})}\right\}\ .
\end{aligned}$$

Solution of $\mathbf{A}_{ij}^{(1)}$

$$\begin{aligned}
\mathbf{A}_{ij}^{(1)} =& \mathbb{E}\left[\Delta Y_i^{(1)} \Delta Y_j^{(2)}\right] \\
=& \mathbb{E}\left[\left(\left|I_i^{(1)}\right|_a \mu_1 + \sigma_1 \int_{t_{i-1}^{(1)}}^{t_i^{(1)}} e^{-\kappa_1\left(t_i^{(1)}-t\right)} \mathrm{d}W_t^{(1)}\right)\left(\left|I_j^{(2)}\right|_b \mu_2 + \sigma_2 \int_{t_{j-1}^{(2)}}^{t_j^{(2)}} e^{-\kappa_2\left(t_j^{(2)}-t\right)} \mathrm{d}W_t^{(2)}\right)\right] \\
=& \mu_1\mu_2 \left|I_i^{(1)}\right|_a \left|I_j^{(2)}\right|_b + \sigma_1\sigma_2 \mathbb{E}\left[\left(\int_{t_{i-1}^{(1)}}^{t_i^{(1)}} e^{-\kappa_1\left(t_i^{(1)}-t\right)} \mathrm{d}W_t^{(1)}\right)\left(\int_{t_{j-1}^{(2)}}^{t_j^{(2)}} e^{-\kappa_2\left(t_j^{(2)}-t\right)} \mathrm{d}W_t^{(2)}\right)\right] .
\end{aligned}$$

Applying Itô isometry we have

$$\begin{aligned}
=& \mu_1\mu_2 \left|I_i^{(1)}\right|_a \left|I_j^{(2)}\right|_b + \sigma_1\sigma_2 \left(\int_{\left|I_i^{(1)} \cap I_j^{(2)}\right|} e^{-\kappa_1\left(t_i^{(1)}-t\right)-\kappa_2\left(t_j^{(2)}-t\right)} \rho_{1,2} \mathrm{d}t\right) \\
=& \mu_1\mu_2 \left|I_i^{(1)}\right|_a \left|I_j^{(2)}\right|_b + \sigma_{1,2} \left|I_i^{(1)} \cap I_j^{(2)}\right|_c .
\end{aligned}$$

Therefore

$$\mathbf{A}_{ij}^{(1)} = \mu_1\mu_2 \left|I_i^{(1)}\right|_a \left|I_j^{(2)}\right|_b + \sigma_{1,2} \left|I_i^{(1)} \cap I_j^{(2)}\right|_c .$$

Solution of $\mathbf{A}_{ij}^{(2)}$ and $\mathbf{A}_{ij}^{(3)}$.
$\mathbf{A}_{ij}^{(2)}$ and $\mathbf{A}_{ij}^{(3)}$ are symmetrical therefore we only solve for $\mathbf{A}_{ij}^{(2)}$.

$$\begin{aligned}
\mathbf{A}_{ij}^{(2)} =& \mathbb{E}\left[\Delta Y_j^{(2)} \sum_h \gamma_h \Delta Y_h^{(1)}\right] \\
=& \mathbb{E}\left[\left(\left|I_j^{(2)}\right|_b \mu_2 + \sigma_2 \int_{I_j^{(2)}} e^{-\kappa_2\left(t_j^{(2)}-t\right)} \mathrm{d}W_t^{(2)}\right)\right. \\
& \left.\left(\int_I \mu_1 \gamma_t e^{-\kappa_1\left(u_t^{(1)}-t\right)} \mathrm{d}t + \sigma_1 \int_I \gamma_t e^{-\kappa_1\left(u_t^{(1)}-t\right)} \mathrm{d}W_t^{(1)}\right)\right] \\
=& \mu_1\mu_2 \left(\int_I \gamma_t a_t \mathrm{d}t\right)\left|I_j^{(2)}\right|_b + \sigma_{1,2} \int_{I \cap I_j^{(2)}} \gamma_t e^{-\kappa_1\left(u_t^{(1)}-t\right)-\kappa_2\left(t_j^{(2)}-t\right)} \mathrm{d}t \\
=& \mu_1\mu_2 \psi \left|I_j^{(2)}\right|_b + \sigma_{1,2} \left|I \cap I_j^{(2)}\right|_c^{\gamma} .
\end{aligned}$$

Therefore

$$\mathbf{A}_{ij}^{(\mathbf{2})} = \mu_1\mu_2\,\psi\left|I_j^{(2)}\right|_b + \sigma_{1,2}\left|I\cap I_j^{(2)}\right|_c^{\gamma}\ .$$

Solution of $\mathbf{A}_{ij}^{(\mathbf{4})}$

$$\begin{aligned}
\mathbf{A}_{ij}^{(\mathbf{4})} =& \mathbb{E}\left[\sum_h \gamma_h \Delta Y_h^{(1)} \sum_k \delta_k \Delta Y_k^{(2)}\right] \\
=& \mathbb{E}\left[\left(\int_I \mu_1\,\gamma_t\,e^{-\kappa_1\left(u_t^{(1)}-t\right)}\mathrm{d}t + \sigma_1\int_I \gamma_t\,e^{-\kappa_1\left(u_t^{(1)}-t\right)}\mathrm{d}W_t^{(1)}\right)\right. \\
& \left.\left(\int_J \mu_2\,\delta_t\,e^{-\kappa_2\left(u_t^{(2)}-t\right)}\mathrm{d}t + \sigma_2\int_J \delta_t\,e^{-\kappa_2\left(u_t^{(2)}-t\right)}\mathrm{d}W_t^{(2)}\right)\right] \\
=& \mu_1\mu_2\left(\int_I \gamma_t\,e^{-\kappa_1\left(u_t^{(1)}-t\right)}\mathrm{d}t\right)\left(\int_J \delta_t\,e^{-\kappa_2\left(u_t^{(2)}-t\right)}\mathrm{d}t\right) \\
& + \sigma_{1,2}\int_{I\cap J} \gamma_t\,\delta_t\,e^{-\kappa_1\left(u_t^{(1)}-t\right)-\kappa_2\left(u_t^{(2)}-t\right)}\mathrm{d}t \\
=& \mu_1\mu_2\,\psi\,\varphi + \sigma_{1,2}\,|I\cap J|_c^{\gamma,\delta}\ .
\end{aligned}$$

Therefore

$$\mathbf{A}_{ij}^{(\mathbf{4})} = \mu_1\mu_2\,\psi\,\varphi + \sigma_{1,2}\,|I\cap J|_c^{\gamma,\delta}\ .$$

Putting all together

$$
\begin{aligned}
\mathbb{E}\left[\Gamma\left(X^{(1)}, X^{(2)}\right)\right] =& \frac{1}{\phi-\eta}\Bigg\{ \sum_{i,j} w_{ij}\ \mathbf{A}_{ij}^{(\mathbf{1})} - \sum_{i,j} w_{ij}\ \alpha_i\ \mathbf{A}_{ij}^{(\mathbf{2})} - \sum_{i,j} w_{ij}\ \beta_j\ \mathbf{A}_{ij}^{(\mathbf{3})} \\
&+ \sum_{i,j} w_{ij}\ \alpha_i\ \beta_j\ \mathbf{A}_{ij}^{(\mathbf{4})} \Bigg\} \\
=& \frac{1}{\phi-\eta}\Bigg\{ \sum_{i,j} w_{ij}\ \left\{ \mu_1\mu_2\ \left|I_i^{(1)}\right|_a \left|I_j^{(2)}\right|_b + \sigma_{1,2}\left|I_i^{(1)} \cap I_j^{(2)}\right|_c \right\} \\
&- \sum_{i,j} w_{ij}\ \alpha_i \left\{ \mu_1\mu_2\,\psi \left|I_j^{(2)}\right|_b + \sigma_{1,2}\left|I \cap I_j^{(2)}\right|_c^{\gamma} \right\} \\
&- \sum_{i,j} w_{ij}\ \beta_j\ \left\{ \mu_1\mu_2\ \left|I_i^{(1)}\right|_a \varphi + \sigma_{1,2}\left|I_i^{(1)} \cap J\right|_c^{\delta} \right\} \\
&+ \sum_{i,j} w_{ij}\ \alpha_i\ \beta_j\ \left\{ \mu_1\mu_2\,\psi\,\varphi + \sigma_{1,2}\left|I \cap J\right|_c^{\gamma,\delta} \right\} \Bigg\} \\
=& \frac{\sigma_{1,2}}{\phi-\eta}\Bigg\{ \sum_{i,j} w_{ij}\ \Bigg\{ \left|I_i^{(1)} \cap I_j^{(2)}\right|_c - \alpha_i \left|I \cap I_j^{(2)}\right|_c^{\gamma} - \beta_j\ \left|I_i^{(1)} \cap J\right|_c^{\delta} \\
&+ \alpha_i\ \beta_j\ \left|I \cap J\right|_c^{\gamma,\delta} \Bigg\} \Bigg\} + \frac{\mu_1\mu_2}{\phi-\eta}\Bigg\{ \sum_{i,j} w_{ij}\ \Big(\left|I_i^{(1)}\right|_a \left|I_j^{(2)}\right|_b \\
&- \alpha_i\ \psi \left|I_j^{(2)}\right|_b - \beta_j\ \left|I_i^{(1)}\right|_a \varphi + \alpha_i\ \beta_j\ \psi\ \varphi \Big) \Bigg\} \\
=& \frac{\sigma_{1,2}}{\phi-\eta}\left\{\mathbf{B}\right\} + \frac{\mu_1\mu_2}{\phi-\eta}\left\{\mathbf{C}\right\}\ .
\end{aligned}
$$

We will no calculate the value $\mathbf{B}$ and $\mathbf{C}$.

Solution of B

$$\begin{aligned}
\mathbf{B} =& \sum_{i,j} w_{ij} \ \left\{ \left| I_i^{(1)} \cap I_j^{(2)} \right|_c - \alpha_i \left| I \cap I_j^{(2)} \right|_c^{\gamma} - \beta_j \ \left| I_i^{(1)} \cap J \right|_c^{\delta} + \alpha_i \ \beta_j \ |I \cap J|_c^{\gamma,\delta} \right\} \\
=& \sum_{i,j} w_{ij} \ \left\{ \left| I_i^{(1)} \cap I_j^{(2)} \right|_c - \frac{\left| I_i^{(1)} \right|_a}{\psi} \left| I \cap I_j^{(2)} \right|_c^{\gamma} - \frac{\left| I_j^{(2)} \right|_b}{\varphi} \ \left| I_i^{(1)} \cap J \right|_c^{\delta} + \frac{\left| I_i^{(1)} \right|_a}{\psi} \ \frac{\left| I_j^{(2)} \right|_b}{\varphi} \ |I \cap J|_c^{\gamma,\delta} \right\} \\
=& \sum_{i,j} w_{ij} \ \left| I_i^{(1)} \cap I_j^{(2)} \right|_c - \sum_{i,j} w_{ij} \ \frac{\left| I_i^{(1)} \right|_a}{\psi} \left| I \cap I_j^{(2)} \right|_c^{\gamma} - \sum_{i,j} w_{ij} \ \frac{\left| I_j^{(2)} \right|_b}{\varphi} \ \left| I_i^{(1)} \cap J \right|_c^{\delta} \\
& + \sum_{i,j} w_{ij} \ \frac{\left| I_i^{(1)} \right|_a}{\psi} \ \frac{\left| I_j^{(2)} \right|_b}{\varphi} \ |I \cap J|_c^{\gamma,\delta} \\
=& \phi - \frac{1}{\psi} \sum_{i,j} w_{ij} \ \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_c^{\gamma} - \frac{1}{\varphi} \sum_{i,j} w_{ij} \ \left| I_j^{(2)} \right|_b \ \left| I_i^{(1)} \right|_c^{\delta} + \frac{|I \cap J|_c^{\gamma,\delta}}{\psi\varphi} \sum_{i,j} w_{ij} \ \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b \\
=& \phi - \left\{ \frac{1}{\psi} \sum_{i,j} w_{ij} \ \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_c^{\gamma} + \frac{1}{\varphi} \sum_{i,j} w_{ij} \ \left| I_j^{(2)} \right|_b \ \left| I_i^{(1)} \right|_c^{\delta} - \frac{|I \cap J|_c^{\gamma,\delta}}{\psi\varphi} \sum_{i,j} w_{ij} \ \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b \right\} .
\end{aligned}$$

Therefore

$$\mathbf{B} = \phi - \eta \ .$$

Solution of C

$$\begin{aligned}
\mathbf{C} =& \sum_{i,j} w_{ij} \ \left(\left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b - \alpha_i \left| I_j^{(2)} \right|_b \psi - \beta_j \ \varphi \left| I_i^{(1)} \right|_a + \alpha_i \ \beta_j \ \psi \ \varphi \right) \\
=& \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b - \sum_{i,j} w_{ij} \ \frac{\left| I^{(1)} *_i \right|_a}{\psi} \ \left| I_j^{(2)} \right|_b \psi \ - \sum_{i,j} w_{ij} \frac{\left| I_j^{(2)} \right|_b}{\varphi} \ \varphi \left| I_i^{(1)} \right|_a \\
& + \sum_{i,j} w_{ij} \frac{\left| I_i^{(1)} \right|_a}{\psi} \ \frac{\left| I_j^{(2)} \right|_b}{\varphi} \ \psi \ \varphi \\
=& \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b - \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b - \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b + \sum_{i,j} w_{ij} \left| I_i^{(1)} \right|_a \left| I_j^{(2)} \right|_b \ .
\end{aligned}$$

Therefore

$$\mathbf{C} = 0 \ .$$

Concluding

$$\begin{aligned}\mathbb{E}\left[\Gamma\left(X^{(1)}, X^{(2)}\right)\right] =& \frac{\sigma_{1,2}}{\phi-\eta}\{\mathbf{B}\} + \frac{\mu_1\mu_2}{\phi-\eta}\{\mathbf{C}\} \\ =& \frac{\sigma_{1,2}}{\phi-\eta}\{\phi-\eta\} + \frac{\mu_1\mu_2}{\phi-\eta}\{0\} \\ =& \sigma_{1,2} \ .\end{aligned}$$

□

Appendix C

Chapter 5

C.1 Details on the setup

We take as given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, and model the death times in a population of n individuals (annuitants or pensioners) as stopping times $\tau^1, \ldots, \tau^n$. This means that at each time t the information carried by $\mathcal{F}_t$ allows us to state whether each individual has died or not. The hedger's liability is given by the random variable $\sum_{i=1}^n 1_{\{\tau^i>T\}}$, which can be equivalently written as $n - \sum_{i=1}^n 1_{\{\tau^i\leq T\}} = n - N_T$. We assume that death times coincide with the first jumps of n conditionally Poisson processes with common random intensity of mortality $(\mu_t)_{t\geq 0}$ under both $\mathbb{P}$ and an equivalent martingale measure $\widetilde{\mathbb{P}}$ [see Biffis et al., 2010, for details]. The expected number of survivors over $[0,T]$ under the two measures can then be expressed as $E^{\mathbb{P}}\left[\sum_{i=1}^n 1_{\{\tau^i>T\}}\right] = np_T$ and $E^{\widetilde{\mathbb{P}}}\left[\sum_{i=1}^n 1_{\{\tau^i>T\}}\right] = n\widetilde{p}_T$, with p_T and $\widetilde{p}_T$ given by the expectation (4.2.1) computed under the relevant probability measure.

Consider any stopping time τ^i satisfying the above assumptions, an integrable random variable $Y \in \mathcal{F}_T$ and a bounded process $(X_t)_{t\in[0,T]}$ such that each X_t is measurable with respect to $\mathcal{F}_{t-}$, the information available up to, but not including, time t. Then a security paying Y at time T in case $\tau^i > T$ and X_{τ^i} at time τ^i in case $\tau^i \leq T$ has time-zero price

$$E^{\widetilde{\mathbb{P}}}\left[\int_0^T \exp\left(-\int_0^s (r_t+\mu_t)\mathrm{d}t\right) X_s\mu_s\mathrm{d}s + \exp\left(-\int_0^T (r_t+\mu_t)\mathrm{d}t\right) Y\right].$$

Consider now two stopping times τ^i, τ^j, with intensities μ^i, μ^j, jointly satisfying the above assumptions (i.e., they are the first jump times of the components of a bivariate conditionally Poisson process). A security paying Y at time T in case neither stopping time has occurred (i.e., $\min(\tau^i, \tau^j) > T$) and X_t in case the first occurrence is at time $t \in (0, T]$ (i.e., $t = \min(\tau^i, \tau^j)$) has time-zero price given by the same formula, with μ_t replaced by $\mu_t^i + \mu_t^j$. This follows from the fact that the stopping time $\min(\tau^i, \tau^j)$ is the first jump time of a conditionally Poisson process with intensity $(\mu_t^i + \mu_t^j)_{t\geq 0}$ [e.g., Bielecki and Rutkowski, 2002]. The expressions presented in sections 4.2-4.4 all follow from these simple results.

Proof of expression (4.4.2). Let $(\delta_t^{\text{H}})_{t\geq 0}$ denote the hedger's net cost of posting collateral and $(\delta_t^{\text{HS}})_{t\geq 0}$ the net yield on the collateral amounts posted by party HS, meaning that holding collateral of amount C_t provides the hedger with an instantaneous yield equal to $\delta_t^{\text{HS}} C_t^+ - \delta_t^{\text{H}} C_t^-$. We assume that collateral is bounded and C_t is measurable with respect to $\mathcal{F}_{t-}$ for all $t \in [0, T]$. Parties H and HS are assumed to have death (default) times satisfying the properties reviewed above, in particular having intensities $\lambda^{\text{H}}, \lambda^{\text{HS}}$. Recalling the recovery rules described in section 4.4, we can then write:

$$\begin{aligned} S_0 =& E^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T (r_t + \lambda_t^{\text{H}} + \lambda_t^{\text{HS}})\mathrm{d}t\right)\left(P - \overline{p}^d\right)\right] \\ &+ E^{\widetilde{\mathbb{P}}}\left[\int_0^T \exp\left(-\int_0^s (r_t + \lambda_t^{\text{H}} + \lambda_t^{\text{HS}})\mathrm{d}t\right)\left(\lambda_s^{\text{H}}(S_s^+ - C_s^-) + \lambda_s^{\text{HS}}(C_s^+ - S_s^-)\right)\mathrm{d}s\right] \\ &+ E^{\widetilde{\mathbb{P}}}\left[\int_0^T \exp\left(-\int_0^s (r_t + \lambda_t^{\text{H}} + \lambda_t^{\text{HS}})\mathrm{d}t\right)(\delta_s^{\text{HS}} C_s^+ - \delta_s^{\text{H}} C_s^-)\mathrm{d}s\right]. \end{aligned} \tag{C.1.1}$$

Using representation (4.4.1), the amount recovered by the nondefaulting counterparty at time $\tau = \min(\tau^{\text{H}}, \tau^{\text{HS}}) \leq T$ is

$$1_{\{\tau=\tau^{\text{H}}\}} S_{\tau-}(c_\tau^{\text{H}} 1_{\{S_{\tau-}<0\}} + 1_{\{S_{\tau-}\geq 0\}}) + 1_{\{\tau=\tau^{\text{HS}}\}} S_{\tau-}(c_\tau^{\text{HS}} 1_{\{S_{\tau-}\geq 0\}} + 1_{\{S_{\tau-}<0\}}), \tag{C.1.2}$$

where we see that $c^{\text{H}}, c^{\text{HS}}$ replace the recovery rates $\psi^{\text{H}}, \psi^{\text{HS}}$ introduced in sec-

tion 4.3. We can then write

$$
\begin{aligned}
S_0 =& E^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T (r_t+\lambda_t^{\mathrm{H}}+\lambda_t^{\mathrm{HS}})\mathrm{d}t\right)\left(P-\overline{p}^d\right)\right] \\
&+ E^{\widetilde{\mathbb{P}}}\left[\int_0^T \exp\left(-\int_0^s (r_t+\lambda_t^{\mathrm{H}}+\lambda_t^{\mathrm{HS}})\mathrm{d}t\right)\left(\lambda_s^{\mathrm{H}}+(\lambda_s^{\mathrm{HS}}+\delta_s^{\mathrm{HS}})c_s^{\mathrm{HS}})S_s^{+}-(\lambda_s^{\mathrm{HS}}+(\lambda_s^{\mathrm{H}}+\delta_s^{\mathrm{H}})c_s^{\mathrm{H}})S_s^{-}\right)\mathrm{d}\right. \\
=& E^{\widetilde{\mathbb{P}}}\left[\exp\left(-\int_0^T (r_t+\Gamma_t)\mathrm{d}t\right)\left(P-\overline{p}^d\right)\right],
\end{aligned}
\tag{C.1.3}
$$

which is nothing other than the usual risk-neutral valuation formula for a security with terminal payoff $S_T = P - \overline{p}^d$ paying continuously a dividend equal to a fraction

$$
(\lambda_s^{\mathrm{H}}+(\lambda_s^{\mathrm{HS}}+\delta_s^{\mathrm{HS}})c_s^{\mathrm{HS}})1_{\{S_{t-}\geq 0\}}+(\lambda_s^{\mathrm{HS}}+(\lambda_s^{\mathrm{H}}+\delta_s^{\mathrm{H}})c_s^{\mathrm{H}})1_{\{S_{t-}<0\}}
$$

of the security's market value an instant before each $t \in [0,T]$. Subtracting the dividend rate from $\lambda^{\mathrm{H}}+\lambda^{\mathrm{HS}}$ and rearranging terms we obtain expression (4.4.2) for Γ.

C.2 Details on the numerical examples

The numerical examples are based on a six-dimensional state variable process $X = (X^{(1)},\ldots,X^{(6)})^{\intercal}$ having $\widetilde{\mathbb{P}}$-dynamics

$$
\begin{aligned}
\mathrm{d}X_t^{(1)} &= \left(k_1(X_t^{(2)}-X_t^{(1)})-\eta_1\right)\mathrm{d}t+\sigma_1\mathrm{d}W_t^{(1)} \\
\mathrm{d}X_t^{(2)} &= \left(k_2(\theta_2-X_t^{(2)})-\eta_2\right)\mathrm{d}t+\sigma_2\mathrm{d}W_t^{(2)} \\
\mathrm{d}X_t^{(3)} &= \left(\kappa_3(\theta_3-X_t^{(3)})+\kappa_{3,1}(X_t^{(1)}-\theta_2)+\kappa_{3,4}(X_t^{(4)}-\theta_4)-\eta_3\right)\mathrm{d}t+\sigma_3\mathrm{d}W_t^{(3)} \\
\mathrm{d}X_t^{(4)} &= \left(\kappa_4(\theta_4-X_t^{(4)})+\kappa_{4,1}(X_t^{(1)}-\theta_2)+\kappa_{4,2}(X_t^{(2)}-\theta_2)-\eta_4\right)\mathrm{d}t+\sigma_4\mathrm{d}W_t^{(4)} \\
\mathrm{d}X_t^{(5)} &= \left(\kappa_5(\theta_5-X_t^{(5)})+\kappa_{5,1}(X_t^{(1)}-\theta_2)+\kappa_{5,2}(X_t^{(2)}-\theta_2)+\kappa_{5,3}(X_t^{(3)}-\theta_3)\right. \\
&\quad \left.+\kappa_{5,4}(X_t^{(4)}-\theta_4)+\kappa_{5,6}(X_t^{(6)}-E_0[X_t^{(6)}])-\eta_5\right)\mathrm{d}t+\sigma_5\mathrm{d}W_t^{(5)} \\
\mathrm{d}X_t^{(6)} &= \left(A_x(t)+B_x(t)(X_t^{(6)}-a_x(t))\right)\mathrm{d}t+\sigma_6(t,x)\mathrm{d}W_t^{(6)},
\end{aligned}
$$

where $W = (W^{(1)}, \ldots, W^{(6)})^\intercal$ is a standard $\widetilde{\mathbb{P}}$-Brownian motion, the constants η_i represent market prices of risk, x is the age of a reference cohort of individuals at time 0, and $A_x(\cdot), B_x(\cdot), \sigma_6(\cdot, x)$ are functions characterizing the dynamics of $X_t^{(6)} = X_{t,x}^{(6)}$ (see below for explicit definitions). The $\mathbb{P}$-dynamics are obtained by removing the market prices of risk from the drifts of the relevant factors and replacing the innovations with the corresponding $\mathbb{P}$-Brownian innovations. We assume that $X^{(6)}$ has the same dynamics under the physical and the pricing probability measures, consistent with our baseline case of a swap rate equal to p_T for each T in the absence of collateral. The Brownian innovations are uncorrelated, with the exception of the pair $(W^{(1)}, W^{(2)})$, whose instantaneous correlation is denoted by $\rho_{1,2}$.

For the first four factors, we use data from Johannes and Sundaresan [2007] who rely on a two-stage maximum likelihood procedure based on weekly data sampled on Wednesdays, from 1990 to 2002, and set the long-run mean of $X^{(3)}$ equal to the average of the 3-month TED spread over the sampling period. For the log-intensity $X^{(6)}$, we use the mortality model described below, and assume that the Brownian component $W^{(6)}$ is uncorrelated with the other ones. The intensity of mortality is modeled using a continuous-time version of the Lee-Carter model [see Biffis et al., 2010]. We first use the annual central death rates $\{m_{y,s}\}$ for US males from the Human Mortality Database to estimate the model $m_{y,s} = \exp(\alpha(y) + \beta(y)K_s)$ for dates $s = 1961, 1962, \ldots, 2007$ and ages $y = 20, 21, \ldots, 89$ with Singular Value Decomposition [see Lee and Carter, 1992]. The resulting estimates for K are then fitted with the process $K_{s+1} = \delta_K K_s + \sigma_K \varepsilon$, with $\varepsilon \sim N(0,1)$. For fixed age $x = 65$, the estimates for $\{\hat{\alpha}(x+h), \hat{\beta}(x+h)\}_{h=0,1,\ldots}$ are interpolated with differentiable functions $a_x(t), b_x(t)$. The functions A_x, B_x, σ_6 are finally obtained by setting $A_x(t) = a_x'(t) + b_x(t)\delta_K$, $B_x(t) = b_x'(t)b_x(t)^{-1}$ and $\sigma_6(t,x) = b_x(t)\sigma_K$. As we consider a single cohort aged x at the reference date 0, here and throughout the paper we simply write $X_t^{(6)} := X_{t,x}^{(6)}$. The extension to multiple (say l) cohorts, would require the analysis of the vector of log-intensities $(X_{t,x_1}^{(6)}, \ldots, X_{t,x_l}^{(6)})$. Although the drift and volatility parameters would be different for each $X_{t,x_i}^{(6)}$, the Lee-Carter specification assumes that all cohorts are affected by the same Brownian component $W^{(6)}$. Other models may instead require the introduction of additional sources of uncertainty.

To estimate the dynamics of $X^{(5)}$, the component of collateral costs related to longevity risk, we implement the procedure discussed in section 4.5, setting the duration $\overline{T}$ of the representative liability equal to 15. We simulate forward all of the other state variables, and at each time step we compute the opportunity cost of capital arising from the capital charges accruing to the hedge supplier based on the simulated mortality and market conditions. The expectation appearing in the drift of $X^{(5)}$ ensures that the longevity capital charges react to departures of realized mortality from the term structure of survival rates estimated at inception. We assume that funding occurs at the LIBOR rate plus a fixed spread of 6%, a conservative value for the cost of internal capital. To obtain the net cost of collateral, we take into account the rebate of the risk-free rate. We estimate the parameters for the dynamics of $X^{(5)}$ based on the simulated realizations of $X^{(5)}$ (an example is depicted in figure 4.7). The parameter estimates are obtained by regressing the simulated dynamics of $X^{(5)}$ on the simulated vector of state variables $(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, X^{(6)})$. We simulate $10,000$ paths over 40 years on a semi-annual grid. For each simulation, we set the parameter θ_5 equal to the average of $X^{(5)}$ along the simulated path. The regression estimates obtained for each simulated path are averaged across all simulations to obtain the final values reported in table 4.2.

References

Y. Aït-Sahalia, J. Fan, and D. Xiu. High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517, 2010. 63

L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multidimensional american options. *Management Science*, pages 1222–1234, 2004. 3, 9, 18, 23, 28, 29

T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003. 33

N. Areal, A. Rodrigues, and M.R. Armada. On improving the least squares monte carlo option valuation method. *Review of Derivatives Research*, 11(1):119–151, 2008. 26

S Assefa, TR Bielecki, S Crépey, and M Jeanblanc. CVA computation for counterparty risk assessment in credit portfolios. In T.R. Bielecki, D. Brigo, and F. Patras, editors, *Recent Advancements in the Theory and Practice of Credit Derivatives*. Bloomberg Press, 2010. 96

F. Audrino and F. Corsi. Realized covariance tick-by-tick in presence of rounded time stamps and general microstructure effects. 2008. 39

A.R. Bacinello, E. Biffis, and P. Millossovich. Pricing life insurance contracts with early exercise features. *Journal of Computational and Applied Mathematics*, 233:27–35, 2009. 99

A.R. Bacinello, E. Biffis, and P. Millossovich. Regression-based algorithms for life insurance contracts with surrender guarantees. *Quantitative Finance*, 10: 1077–1090, 2010. 99, 116

F.M. Bandi and J.R. Russell. Realized covariation, realized beta and microstructure noise. *Unpublished paper, Graduate School of Business, University of Chicago*, 2005. 63

O.E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001. 75

O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004. 33

O.E. Barndorff-Nielsen, P.R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169, 2011. 63

E. Barucci and R. Reno. On measuring volatility of diffusion processes with high frequency data. *Economics Letters*, 74(3):371–378, 2002. 38

Basel II. *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Basel Committee on Banking Supervision, 2006. 95

D. Bauer, F.E. Benth, and R. Kiesel. Modeling the forward surface of mortality. Technical report, Georgia State University, 2010a. 100

D. Bauer, D. Bergmann, and A. Reuss. Solvency II and nested simulations - A Least-Squares Monte Carlo approach. Technical report, University of Ulm and Georgia State University, 2010b. 99

D. Bauer, F.E. Benth, and R. Kiesel. Modeling the forward surface of mortality. Technical report, Georgia State University, forthcoming in *SIAM Journal of Financial Mathematics*, 2012. 100

D. Belomestny, C. Bender, and J. Schoenmakers. True upper bounds for bermudan products via non-nested monte carlo. *Mathematical Finance*, 19(1):53–71, 2009. 1, 3, 9, 11, 13, 14, 18, 25, 26, 29, 30, 31

M. Bibinger. Efficient covariance estimation for asynchronous noisy high-frequency data. *Scandinavian Journal of Statistics*, 38(1):23–45, 2011. 39

T. Bielecki and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer, 2002. 113, 141

E. Biffis and D. Blake. Informed intermediation of longevity exposures. Technical report, CAREFIN, Bocconi Milan, 2009. 100

E. Biffis and D. Blake. Securitizing and tranching longevity exposures. *Insurance: Mathematics & Economics*, 46(1):186–197, 2010a. 100

E. Biffis and D. Blake. Mortality-linked securities and derivatives. In M. Bertocchi, S.L. Schwartz, and W.T. Ziemba, editors, *Optimizing the Aging, Retirement and Pensions Dilemma*. John Wiley & Sons, 2010b. 104

E. Biffis, M. Denuit, and P. Devolder. Stochastic mortality under measure changes. *Scandinavian Actuarial Journal*, 2010:284–311, 2010. 100, 102, 140, 143

E. Biffis, D. Blake, L. Pitotti, and A. Sun. The cost of counterparty risk and collateralization in longevity swaps. *Available at SSRN 1801826*, 2012. 93, 94

Å. Björck. Solving linear least squares problems by gram-schmidt orthogonalization. *BIT Numerical Mathematics*, 7(1):1–21, 1967. 24

F. Black and M. Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973. 19

T. Bollerslev and B.Y.B. Zhang. Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, 10 (5):533–558, 2003. 63

P.P. Boyle. Option valuation using a three-jump process. *International Options Journal*, 3(7-12):1, 1986. 2

M.J. Brennan and E.S. Schwartz. The valuation of american put options. *The Journal of Finance*, 32(2):449–462, 1977. 2

D. Brigo and A. Capponi. Bilateral counterparty risk valuation with stochastic dynamical models and application to CDSs. Technical report, King's College London and California Institute of Technology, 2009. 96

D. Brigo, A. Pallavicini, and V. Papatheodorou. Arbitrage-free valuation of bilateral counterparty risk for interest-rate products: Impact of volatilities and correlations. *International Journal of Theoretical and Applied Finance*, 14(6): 773802, 2011. 96

D. Brigo, A. Capponi, and A. Pallavicini. Arbitrage-free bilateral counterparty risk valuation under collateralization and application to credit default swaps. *Mathematical Finance*, 2012. to appear, `http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9965.2012.00520.x/pdf`. 96

A.J.G. Cairns, D. Blake, K.G. Dowd, G.D. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, 13(1):1–35, 2009. 102

A.J.G. Cairns, D. Blake, K.G. Dowd, G.D. Coughlan, D. Epstein, and M. Khalaf-Allah. Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics & Economics*, 48:355–367, 2011. 106

P. Carr, H. Geman, D.B. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation*. *The Journal of Business*, 75(2):305–333, 2002. 19

J.F. Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: mathematics and Economics*, 19(1): 19–30, 1996. 3, 6

CEIOPS. SCR standard formula: Further advice on the counterparty default risk module. Technical report, Committee of European Insurance and Occupational Pension Supervisors' Consultation Paper no. 51, 2009. 95

H. Chen and J.D. Cummins. Longevity bond premiums: The extreme value approach and risk cubic pricing. *Insurance: Mathematics & Economics*, 46(1): 150–161, 2010. 100

E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression method for american option pricing. *Finance and Stochastics*, 6(4): 449–471, 2002. 3, 8

K.J. Cohen, G.A. Hawawini, S.F. Maier, R.A. Schwartz, and D.K. Whitcomb. Friction in the trading process and the estimation of systematic risk. *Journal of Financial Economics*, 12(2):263–278, 1983. 63

R. Cont and P. Tankov. *Financial modelling with jump processes*, volume 2. Chapman & Hall, 2004. 13

F. Corsi, S. Peluso, and F. Audrino. Missing in asynchronicity: A kalman-em approach for multivariate realized covariance estimation. 2012. 36, 61

G.D. Coughlan, M. Khalaf-Allah, Y. Ye, S. Kumar, A.J.G. Cairns, D. Blake, and K. Dowd. Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. Technical report, to appear in *North American Actuarial Journal*, 2011. 104

J.C. Cox, S.A. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of financial Economics*, 7(3):229–263, 1979. 2

S.H. Cox, Y. Lin, and H. Pedersen. Mortality risk modeling: Applications to insurance securitization. *Insurance: Mathematics & Economics*, 46(1):242–253, 2010. 100

J.D. Cummins and P. Trainar. Securitization, insurance, and reinsurance. *Journal of Risk and Insurance*, 76(3):463–492, 2009. 95

M.M. Dacorogna. *An introduction to high-frequency finance*. Academic Pr, 2001. 38

MHA Davis and I. Karatzas. A deterministic approach to optimal stopping. *Probability, Statistics and Optimisation (ed. FP Kelly). NewYork Chichester: John Wiley & Sons Ltd*, pages 455–466, 1994. 3, 8, 12

F. De Jong and T. Nijman. High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance*, 4(2):259–277, 1997. 63

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977. 43, 78

K. Dowd, D. Blake, A.J.G. Cairns, and P. Dawson. Survivor swaps. *Journal of Risk and Insurance*, 73(1):1–17, 2006. 94, 100

K. Dowd, A.J.G. Cairns, D. Blake, G.D. Coughlan, D. Epstein, and M. Khalaf-Allah. Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. *North American Actuarial Journal*, 14(3):281–298, 2010a. 106

K. Dowd, A.J.G. Cairns, D. Blake, G.D. Coughlan, D. Epstein, and M. Khalaf-Allah. Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics & Economics*, 47:255–265, 2010b. 106

D. Duffie and M. Huang. Swap rates and credit quality. *Journal of Finance*, 51 (3):921–949, 1996. 109, 116

T.W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, pages 291–298, 1979. 34

G. E. Forsythe, M.A. Malcolm, and C.B. Moler. *Computer Methods for Mathematical Computations*. Prentice-Hall, 1976. 117

S. Gerhold. The longstaff–schwartz algorithm for lévy models: Results on fast and slow convergence. *The Annals of Applied Probability*, 21(2):589–608, 2011. 19

F. Girosi and G. King. *Demographic Forecasting*. Princeton University Press, 2008. 106

P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer verlag, 2004. 29

P. Glassermann. *Monte Carlo Methods in Financial Engineering*. Springer, 2004. 99

W.H. Greene. *Econometric Analysis*. Pearson/Prentice Hall, Upper Saddle River, NJ, sixth edition, 2008. 35

J.E. Griffin and R.C.A. Oomen. Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*, 160(1):58–68, 2011. 39, 40, 83

P. Hansen and A. Lunde. An unbiased measure of realized variance. 2004. 38

M.B. Haugh and L. Kogan. Pricing american options: a duality approach. *Operations Research*, pages 258–270, 2004. 3, 8, 9

T. Hayashi and S. Kusuoka. Consistent estimation of covariation under nonsynchronicity. *Statistical Inference for Stochastic Processes*, 11(1):93–106, 2008. 34

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005. 32, 34, 35, 38, 39, 54, 59, 60, 63, 70, 87, 92

T. Hayashi and N. Yoshida. Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics*, 60(2):367–406, 2008. 34, 35

M. Hörig and M. Leitschkis. Solvency ii proxy modelling via least squares monte carlo. Technical report, Milliman, http://publications.milliman.com/publications/life-published/pdfs/solvency-II-p 2012. 99

J. Inkmann and D. Blake. Managing financially distressed pension plans in the interest of beneficiaries. Technical report, Pensions Institute Discusson Paper PI-0709, 2010. 97

ISDA. *Credit Support Annex*. International Swaps and Derivatives Association, 1994. 95

ISDA. *Big Bang Protocol*. International Swaps and Derivatives Association, 2009. 96

ISDA. *Market Review of OTC Derivative Bilateral Collateralization Practices*. International Swaps and Derivatives Association, 2010a. 97, 111, 114

ISDA. *2010 Margin Survey*. International Swaps and Derivatives Association, 2010b. 95, 97

F. Jamshidian. Chaotic expansion of powers and martingale representation. 2005. 11, 12

M. Johannes and S. Sundaresan. The impact of collateralization on swap rates. *Journal of Finance*, 62(1):383–410, 2007. 97, 98, 111, 114, 116, 118, 119, 143

T. Kanatani. Integrated volatility measuring from unevenly sampled observations. *Economics Bulletin*, 3(36):1–8, 2004. 38

I. Karatzas and S.E. Shreve. *Methods of mathematical finance*, volume 39. Springer Verlag, 1998. 5

SG Kou. A jump-diffusion model for option pricing. *Management Science*, pages 1086–1101, 2002. 19

D. Lakdawalla and G. Zanjani. Catastrophe bonds, reinsurance, and the optimal collateralization of risk transfer. Technical report, RAND Corporation and Federal Reserve Bank of New York, 2007. 95

R.D. Lee and L.R. Carter. Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671, 1992. 102, 143

F.A. Longstaff and E.S. Schwartz. Valuing american options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001. 2, 6, 9, 17

M. Ludkovski and V.R. Young. Indifference pricing of pure endowments and life annuities under stochastic hazard and interest rates. *Insurance: Mathematics and Economics*, 42(1):14–30, 2008. 100

D.B. Madan and F. Milne. Option pricing with vg martingale components. *Mathematical Finance*, 1(4):39–55, 1991. 26

D.B. Madan and E. Seneta. The variance gamma (vg) model for share market returns. *Journal of business*, pages 511–524, 1990. 19, 26

D.B. Madan, P.P. Carr, and E.C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1):79–105, 1998. 27

M.E. Mancino and S. Sanfelici. Estimating covariance via fourier method in the presence of asynchronous trading and microstructure noise. *Journal of Financial Econometrics*, 9(2):367–408, 2011. 63

M. Martens. Estimating unbiased and precise realized covariances. In *EFA 2004 Maastricht Meetings Paper No. 4299*, 2004. 34

R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1):125–144, 1976. 19

S.C. Narula. Orthogonal polynomial regression. *International Statistical Review/Revue Internationale de Statistique*, pages 31–36, 1979. 23

D. Nualart and W. Schoutens. Chaotic and predictable representations for lévy processes. *Stochastic Processes and their Applications*, 90(1):109–122, 2000. 1, 3, 11, 13, 15, 31

A. Palandri. Consistent realized covariance for asynchronous observations contaminated by market microstructure noise. *Unpublished manuscript*, 2006. 39, 63

E. Pitacco, M. Denuit, S. Haberman, and A. Olivieri. *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press, 2009. 106

F. Rapisarda, D. Brigo, and F. Mercurio. Parameterizing correlations: a geometric interpretation. *IMA Journal of Management Mathematics*, 18(1):55–73, 2007. 65

R. Rebonato and P. Jackel. The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk*, 2:17–28, 2000. 62, 64, 65, 71, 90, 92

R. Renò. A closer look at the epps effect. *International Journal of Theoretical and Applied Finance*, 6(1):87–102, 2003. 34, 63

J.R. Rice. Experiments on gram-schmidt orthogonalization. *Mathematics of Computation*, 20(94):325–328, 1966. 23

L.C.G. Rogers. Monte carlo valuation of american options. *Mathematical Finance*, 12(3):271–286, 2002. 3, 8, 14

Y. Salhi and S. Loisel. Joint modeling of portfolio experienced and national mortality: A co-integration based approach. Technical report, ISFA Lyon, 2010. 104

J. Schoenmakers, J. Huang, and J. Zhang. Optimal dual martingales, their analysis and application to new algorithms for bermudan products. *Arxiv preprint arXiv:1111.6038*, 2012. 3, 9, 11, 29, 30, 31

M. Scholes and J. Williams. Estimating betas from nonsynchronous data. *Journal of financial economics*, 5(3):309–327, 1977. 63

N. Shephard and D. Xiu. Econometric analysis of multivariate realised qml: efficient positive semi-definite estimators of the covariation of equity prices. 2012. 36, 42, 61, 74, 83

K. Sheppard. Realized covariance and scrambling. *Unpublished manuscript*, 2006. 63

S.E. Shreve. *Stochastic calculus for finance: Continuous-time models*, volume 2. Springer Verlag, 2004. 4, 5

R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis*, 3(4):253–264, 1982. 36

L. Stentoft. Assessing the least squares monte-carlo approach to american option valuation. *Review of Derivatives research*, 7(2):129–168, 2004a. 7

L. Stentoft. Convergence of the least squares monte carlo approach to american option valuation. *Management Science*, pages 1193–1203, 2004b. 3, 8

R.A. Stevens, A. De Waegenaere, and B. Melenberg. Calculating capital requirements for longevity risk in life insurance products: Using an internal model in line with Solvency II. Technical report, Tilburg University, 2010a. 122

R.A. Stevens, A. De Waegenaere, and B. Melenberg. Longevity risk and hedge effects in a portfolio of life insurance products with investment risk. Technical report, Tilburg University, 2010b. 104

The Actuarial Profession. *Allowing for the Sponsor Covenant in Actuarial Advice*. Sponsor Covenant Working Party Final Report, 2005. 108

The Pensions Regulator. *Scheme Funding and the Employer Covenant*. www.thepensionsregulator.gov.uk/docs/employer-covenant-statement-june-2009.pdf 2009. 108

J.N. Tsitsiklis and B. Van Roy. Optimal stopping of markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-

dimensional financial derivatives. *Automatic Control, IEEE Transactions on*, 44(10):1840–1851, 1999. 3, 6

J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american-style options. *Neural Networks, IEEE Transactions on*, 12(4):694–703, 2001. 3, 6

V. Voev and A. Lunde. Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics*, 5(1):68–104, 2007. 39

D. Williams. *Probability with martingales*. Cambridge Univ Pr, 1991. 8

D. Xiu. Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159(1):235–250, 2010. 42, 74, 83, 84

L. Zhang. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47, 2011. 34, 38, 63