New Calibration Methods for Continuous-Time Stochastic Models in Finance, Part I:

Maximum Likelihood Methods for 1-Dimensional Models with Open-Close and Open-High-Low-Close Transition Densities Approximated via Heat Kernels

Dr Leonid Merkin-Janson

27 Jan 2019 (6th Revision)

Abstract

For a 1-dimensional stochastic differential equation of the type $dS_t = \tilde{\mu}(S_t, t; \boldsymbol{q})dt + \tilde{\sigma}(S_t, t; \boldsymbol{q})dW_t$, where S_t is e.g. the price of a certain financial asset, W_t a Brownian motion and $\boldsymbol{q}$ is a finite-dimensional vector of unknown model parameters, we propose a new method for estimating the vector $\boldsymbol{q}$ on the historical time series of S_t. Our method is based on the classical Maximum Likelihood Estimation (MLE) framework, but differs from the previously-known MLE methods in two major respects. *First,* the likelihood values are computed using probability density functions (PDFs) for 3-variate distributions of Open, High, Low and Close (OHLC) prices for each historical interval, instead of commonly-used 1-variate PDFs of Open and Close prices only. This allows us to take into account additional market information, i.e. *ranges* of S_t values as opposed to just point-wise values. *Secondly*, in order to construct the required 3-variate PDFs, we enhance a previously-known formula which was applicable to constant coefficients $\tilde{\mu}$ and $\tilde{\sigma}$ only (i.e., S_t being a Brownian motion with drift). By combining that formula with results from Heat Kernel theory, we extend it to the case of non-constant $\tilde{\mu}$ and $\tilde{\sigma}$, thus making our method applicable to a wide class of realistic stochastic models in Quantitative Finance.

Contents

1 Maximum Likelihood Estimation Framework

Consider a 1-dimensional stochastic differential equation (1D SDE for short)

$$dS_t = \tilde{\mu}(S_t, t; \boldsymbol{q})\, dt + \tilde{\sigma}(S_t, t; \boldsymbol{q})\, dW_t \ , \tag{1}$$

where S_t is a stochastic process representing e.g. the price of a certain financial asset or instrument, t is time, $\tilde{\mu}(S_t, t; \boldsymbol{q})$ is called a *trend (drift)* term, $\tilde{\sigma}(S_t, t; \boldsymbol{q})$ is a called a *volatility* term, $\boldsymbol{q} \in \mathbb{R}^K$ is a vector of model parameters, and W_t is a Brownian motion. For simplicity, we will assume that $\tilde{\mu}$ and $\tilde{\sigma}$ are continuously-differentiable functions of all their arguments, and also

$$\tilde{\sigma}(S_t, t; \boldsymbol{q}) > 0 \tag{2}$$

with probability 1.

Typically, we can make some modeling assumptions about the functional structure of drift and volatility terms in (1), but the vector of parameters $\boldsymbol{q}$ is a priori unknown, and needs to be estimated (calibrated) on historical time series (market data) of S_t.

To that end, suppose that we got a sequence of time instants $t_0 < t_1 < \ldots < t_N$ and the corresponding values of S_t: $S_{t_i} = S_i$, $0 \le i \le N$. The classical **Maximum Likelihood Estimation (MLE)** framework of methods for obtaining the vector of parameters $\boldsymbol{q}$ can be outlined as follows.

To begin with, since the process S_t defined by SDE (1) is a 1D *Markovian process* [14], the probability distribution of S_t is fully determined by its "last known" value and the parameters $\boldsymbol{q}$. Thus, consider a time interval (t_{i-1}, t_i) $(1 \le i \le N)$ and the probability density function (PDF) of S_t in this interval, conditioned on the value of S_t at the beginning of the interval, i.e. S_{i-1}. This conditional PDF will be denoted by

$$p_{i-1}(s, t; \boldsymbol{q}) = p(s, t; \boldsymbol{q} \,|\, S_{i-1}) \ , \quad t_{i-1} \le t < t_i \ . \tag{3}$$

According to the Markovian property, $p_{i-1}(s, t; \boldsymbol{q})$ is fully determined by the values of S_{i-1}, t_{i-1} and the parameter vector $\boldsymbol{q}$. It does not depend on any previous values $S_0, \ldots, S_{i-2}$ (and needless to say, it cannot depend on any future values $S_i, \ldots, S_N$ either). Throughout this paper, we will use the term *"transition density"* for PDF (3). As a consequence of the Markovian property, the transition densities for different intervals (t_{i-1}, t_i) are independent from each other.

Therefore, the likelihood of S_t making a transition from a given point S_{i-1} at time t_{i-1} to any point in the infinitesimal interval $(S_i, S_i + dS_i)$ at time t_i, is given by

$$dP_i(\boldsymbol{q}) = \mathrm{Prob}\{S_i \le S_{t_i} \le S_i + dS_i \,|\, S_{i-1}\} \ = \ p_{i-1}(S_i, t_i; \boldsymbol{q})\, dS_i \ . \tag{4}$$

(Note that we have to consider an infinitesimal neighborhood of S_i rather than a fixed target point S_i, because the likelihood of an exact point-to-point transition $S_{i-1} \mapsto S_i$ would always be 0).

Then by the independence property mentioned above, the likelihood of S_t following a path from a fixed point S_0 at t_0 through the neighborhoods $(S_i, S_i + dS_i)$ $(1 \le i \le N)$ at the corresponding time instants t_i, would be given by

$$\prod_{i=1}^{N} p_{i-1}(S_i, t_i; \boldsymbol{q})\, dS_i \ ,$$

up to higher-order terms (of the order $o(dS_1 \ldots dS_N)$).

The main idea of the MLE framework is to set all neighborhood sizes to a constant value (which may be arbitrarily small but fixed from now on):

$$dS_1 = \ldots = dS_N = \delta = \text{ const } > 0$$

and to select the vector of parameters $\boldsymbol{q}^*$ which maximizes the "likelihood density" function

$$\tilde{L}_1(\boldsymbol{q}) = \prod_{i=1}^{N} p_{i-1}(S_i, t_i; \boldsymbol{q}) \ , \tag{5}$$

over the domain $\mathcal{Q}$ of admissible values of $\boldsymbol{q}$ (which depends on a particular model (1)):

$$\boldsymbol{q}^* = \arg\max_{\boldsymbol{q}\in\mathcal{Q}} \tilde{L}_1(\boldsymbol{q}) \ .$$

From the computational point of view, it is usually more convenient to maximize $\log \tilde{L}_1(\boldsymbol{q})$ instead of $\tilde{L}_1(\boldsymbol{q})$ itself:

$$\begin{aligned} L_1(\boldsymbol{q}) &= \log \tilde{L}_1(\boldsymbol{q}) = \sum_{i=1}^{N} \log\left(p_{i-1}(S_i, t_i; \boldsymbol{q})\right) \ , \\ \boldsymbol{q}^* &= \arg\max_{\boldsymbol{q}\in\mathcal{Q}} L_1(\boldsymbol{q}) \ . \end{aligned} \tag{6}$$

This concludes the general formulation of the standard MLE framework.

2 Challenges of Implementing MLE Methods

In practice, implementation of any particular MLE method in Quantitative Finance requires the following problems to be addressed:

Computation of transition densities (3): For some simple models, the transition PDFs may be known in a closed analytical form (e.g. when they are normal or log-normal; see Section 6 for an example). However, in the general case of (1), exact expressions for transition densities are not known, so various approximation schemes can be proposed. This problem is discussed in more detail in Section 3, including our innovative Heat Kernel-based approach [2, 12] to constructing transition densities.

An alternative approach to handling the situations when transition densities are not available, is to abandon the MLE framework altogether and to replace it with a more general but arguably less precise *Minimum Contrast Principle* [9]. In our case, we can stay within the MLE framework owing to the Heat Kernel methods which allow us to construct sufficiently accurate 1-variate and 3-variate (see below) transition densities.

Optimal intervalization of historical market data: On one hand, using smaller time intervals $\tau_i = t_i - t_{i-1}$ would generally be a good idea, as it would make more data points available for model calibration (for the same over-all calibration interval $T = t_N - t_0$). On the other hand, one should bear in mind that a model given by (1) is based on certain fundamental assumptions; in particular, the price process S_t in (1) is driven by the Brownian motion W_t which is a continuous stochastic process with zero auto-correlation of increments. For very short time intervals, those assumptions will become invalid due to the effects of market micro-structure.

Thus, there exists a reasonable lower limit for time intervals τ_i. A *"rule of thumb"* is that for top-liquid financial instruments, τ_i should be at least 1 minute. A still better

approach would be use a "liquidity clock" [13]. In that case, t is re-defined: it is no longer a wall-clock time, but a measure of the amount of contracts traded. For the purpose of our subsequent analysis, the exact definition of t is irrelevant, so we will continue to call it "time".

Using OHLC market data: The above considerations regarding the lower bound on intervals τ_i give rise to a problem of how to use *all* available market information efficiently in MLE calibration, without going for too short time intervals.

Specifically, for each interval (t_{i-1}, t_i), in addition to the *"Open"* price S_{i-1} and the *"Close"* price S_i, we would like to incorporate into our analysis the *"High"* price H_i and the *"Low"* price L_i (*"OHLC prices"* for short). Here

$$L_i = \min_{t_{i-1} \le t \le t_i} S_t \ , \quad H_t = \max_{t_{i-1} \le t \le t_i} S_t \ .$$

To that end, let us consider a joint 3-dimensional stochastic process (S_t, L_t, H_t), where L_t and H_t are "low" and "high" (resp.) values of S_t since some fixed starting time t_* and up to time t:

$$L_t = \min_{t_* \le t' \le t} S_{t'} \ , \quad H_t = \max_{t_* \le t' \le t} S_{t'} \ . \tag{7}$$

It is possible to show that this joint process is also Markovian [14]; its 3-dimensional probability distributions within each interval (t_{i-1}, t_i), conditioned on the corresponding initial triple $(S_{i-1}, S_{i-1}, S_{i-1})$ at starting time $t_* = t_{i-1}$, are independent from one interval to another, and are determined only by S_{i-1} and the parameter vector $\boldsymbol{q}$. Obviously, we get $L_i = L_{t_i}$ and $H_i = H_{t_i}$.

Let

$$p_{i-1}(s, l, h, t; \boldsymbol{q}) = p(s, l, h, t; \boldsymbol{q} \,|\, S_{i-1}) \tag{8}$$

be a "3-variate" PDF of the joint process (S_t, L_t, H_t) within an interval (t_{i-1}, t_i), conditional on S_{i-1}. (Note that by terminological convention, the number of "variates" does not include t, condition variables or model parameters). Due to the Markovian property pointed out above, we can use this 3-variate transition PDF in construction of the likelihood of the process (S_t, L_t, H_t) passing through the neighborhoods of the observed values $\{(S_i, L_i, H_i)\}$ at the corresponding times $\{t_i\}$, $1 \le i \le N$:

$$\prod_{i=1}^{N} p_{i-1}(S_i, L_i, H_i, t_i)\, dS_i\, dL_i\, dH_i \ .$$

Similar to the "1-variate" case, we then set all neighborhood sizes to an arbitrary fixed value δ and obtain the vector of parameters $\boldsymbol{q}^*$ by maximizing the following "log-likelihood density":

$$\begin{aligned} L_3(\boldsymbol{q}) &= \sum_{i=1}^{N} \log\left(p_{i-1}(S_i, H_i, L_i, t_i; \boldsymbol{q})\right) \ , \\ \boldsymbol{q}^* &= \arg\max_{\boldsymbol{q} \in \mathcal{Q}} L_3(\boldsymbol{q}) \ . \end{aligned} \tag{9}$$

(Needless to say, although the values L_i and H_i are formally attributed to the time instant t_i, it does not mean that the corresponding low or high value of S_t had occurred exactly at time t_i; we can only state that they occurred at some time instants within the interval (t_{i-1}, t_i)).

The methods for computing 3-variate transition densities and constructing the MLE calibration procedures based on them, will be discussed in detail in Section 4.

Numerical optimization methods: Typically, the log-likelihood functions (6) or (9) cannot be maximized analytically; a numerical method for constrained global optimization is required to that end. However, the problems related to numerical optimization are beyond the scope of this paper.

Remark:
In general, 3-variate transition PDFs are quite difficult to compute, and to the best of our knowledge, have not yet been used in MLE methods. There are multiple well-known range-based volatility estimators [11] which involve Open, High, Low and Close prices, but those estimators differ from our approach in several respects:

- The existing estimators are model-free and non-parametric, and for this reason they do not rely on any explicit transition PDFs; whereas we follow a model-driven parametric approach which stems from a particular model of type (1), and use explicit methods for construction of transition densities (e.g. Fokker-Planck PDE, Heat Kernels) as described in Sections 3 and 4.
- Furthermore, the existing estimators are intended to obtain only the volatility values; whereas our approach is aimed at estimating the full vector of parameters $\boldsymbol{q}$, including those which belong to the trend (drift) term.

■

3 Computation of 1-Variate Transition Densities

3.1 Locally-Normal Approximation

Let us return to SDE (1) and consider the stochastic process S_t within an interval $t_{i-1} \le t < t_i$, conditioned on the initial value $S_{t_{i-1}} = S_{i-1}$. We can assume that the time differences

$$\tau = t - t_{i-1}$$

are sufficiently small (though not infinitesimal, see Section 2). In that case, we can approximate the SDE by fixing $s = S_{i-1}$, $t = t_{i-1}$ in the trend and volatility terms:

$$\begin{aligned} \tilde{\mu}_{i-1}(\boldsymbol{q}) &= \tilde{\mu}(S_{i-1}, t_{i-1}; \boldsymbol{q}) \ , \\ \tilde{\sigma}_{i-1}(\boldsymbol{q}) &= \tilde{\sigma}(S_{i-1}, t_{i-1}; \boldsymbol{q}) \ , \\ dS_t &\simeq \tilde{\mu}_{i-1}(\boldsymbol{q})\, dt + \tilde{\sigma}_{i-1}(\boldsymbol{q})\, dW_t \ . \end{aligned} \tag{10}$$

SDE (10) has constant coefficients, so it can easily be solved. The resulting distribution of S_t will be a locally-normal one, and we get a normal 1-variate transition density for S_t, to be inserted into (6):

$$\begin{aligned} S_t &= S_{i-1} + \tilde{\mu}_{i-1}(\boldsymbol{q})\, \tau + \tilde{\sigma}_{i-1}(\boldsymbol{q})\, W_\tau \ , \\ S_t &\sim \mathcal{N}\left(S_{i-1} + \tilde{\mu}_{i-1}(\boldsymbol{q})\, \tau,\ \tilde{\sigma}_{i-1}(\boldsymbol{q})\sqrt{\tau}\right) \ , \\ p_{i-1}(s, t; \boldsymbol{q}) &= \frac{1}{\sqrt{2\pi\tau}\tilde{\sigma}_{i-1}(\boldsymbol{q})} \exp\left\{-\frac{1}{2\tau}\left(\frac{s - S_{i-1} - \tilde{\mu}_{i-1}(\boldsymbol{q})\tau}{\tilde{\sigma}_{i-1}(\boldsymbol{q})}\right)^2\right\} \ . \end{aligned} \tag{11}$$

It is important to note that even under this approximation, the distribution of S_t is only *locally-normal*: it is considered to be normal within each interval (t_{i-1}, t_i), but the corresponding

normal expectation and volatility would in general differ from one interval to another, so the "global" distribution of S_t would not be normal.

After trivial transformations, the resulting estimator will be the following:

$$\begin{aligned} L_1^{\mathrm{LN}}(\boldsymbol{q}) &= \sum_{i=1}^{N} \left[\log \tilde{\sigma}_{i-1}(\boldsymbol{q}) + \frac{1}{2\tau_i} \left(\frac{S_i - S_{i-1} - \tilde{\mu}_{i-1}(\boldsymbol{q})\,\tau_i}{\tilde{\sigma}_{i-1}(\boldsymbol{q})} \right)^2 \right] , \\ \boldsymbol{q}^* &= \arg\min_{\boldsymbol{q}\in\mathcal{Q}} L_1^{\mathrm{LN}}(\boldsymbol{q}) , \end{aligned}$$

where $\tau_i = t_i - t_{i-1}$ (same notation as in Section 2).

In practice, the locally-normal approximation (10) may or may not be suitable. In many cases, it actually appears to be quite crude. For example:

- For some asset classes (e.g. Equities, FX, Bonds), the price dynamics of S_t could be much more closely approximated by a *log-normal distribution*, yet there is no provision for a log-transform in (10).
- In many cases, the actual distribution of S_t would feature *"heavy tails"* (that is, the probabilities of large S_t movements over a given time interval would greatly exceed the corresponding normal probabilities). If we restrict our attention to 1D models (1), the primary origin of "heavy tails" is in the structure of the volatility term $\tilde{\sigma}(s,t;\boldsymbol{q})$, because for infinitesimal dt, $\tilde{\sigma}\,dW_t$ is of order $O(\sqrt{dt})$, whereas $\tilde{\mu}\,dt$ is of higher order $O(dt)$. Thus, in such cases, a locally-constant approximation (10) for $\tilde{\sigma}$ would be particularly unsuitable.

3.2 Tail Normalization Transform

3.2.1 Normalizing the Volatility to 1

The first step to rectify the shortcomings of the locally-normal approximation (10) is to apply a non-linear transform to S_t which will make the volatility term of the transformed process to be *exactly constant*, e.g. unitary [1].

Let us introduce a transform $f : s \to x$ parameterized by t and $\boldsymbol{q}$:

$$x = f(s;t,\boldsymbol{q}) = \int_{S_*}^{s} \frac{dS}{\tilde{\sigma}(S,t;\boldsymbol{q})} , \tag{12}$$

where S_* is some fixed integration limit. Note that the integrand in (12) is finite and positive with probability 1, due to condition (2). Thus, for fixed t and $\boldsymbol{q}$, f is a monotonically-increasing function of s. Let us also denote by $f^{-1}(x;t,\boldsymbol{q})$ the inverse map $x \to s$ taken under the same parameters t and $\boldsymbol{q}$ as f.

Then consider the stochastic process

$$X_t = f(S_t;t,\boldsymbol{q}) . \tag{13}$$

By applying the Itô–Döblin formula [14], we get the following SDE for X_t, with volatility being equal to 1 exactly:

$$\begin{aligned} dX_t &= \frac{\partial f}{\partial t}\,dt + \frac{\partial f}{\partial s}\,dS_t + \frac{1}{2}\frac{\partial^2 f}{\partial s^2}\,dS_t^2 = \hat{\beta}(X_t,t;\boldsymbol{q})\,dt + dW_t , \\ &\text{where} \\ \hat{\beta}(X_t,t;\boldsymbol{q}) &= \tilde{\beta}(f^{-1}(X_t;t,\boldsymbol{q}),t;\boldsymbol{q}) , \\ \tilde{\beta}(s,t;\boldsymbol{q}) &= \frac{\tilde{\mu}(s,t;\boldsymbol{q})}{\tilde{\sigma}(s,t;\boldsymbol{q})} - \frac{1}{2}\frac{\partial\tilde{\sigma}}{\partial s}(s,t;\boldsymbol{q}) - \int_{S_*}^{s} \frac{\partial\tilde{\sigma}}{\partial t}(S,t;\boldsymbol{q})\,\frac{dS}{\tilde{\sigma}^2(S,t;\boldsymbol{q})} . \end{aligned} \tag{14}$$

This transform (called "Tail Normalization" in [1]) addresses both problems pointed out in Section 3.1:

- For a log-normal process given by $\tilde{\mu} = \mu S_t$ and $\tilde{\sigma} = \sigma S_t$, where $\boldsymbol{q} = [\mu, \sigma]$ is a vector of constant parameters, $\sigma > 0$, we would get $X_t = \frac{1}{\sigma} \log \frac{S_t}{S_*}$ and $\hat{\beta} = \frac{1}{\sigma}\left(\mu - \frac{\sigma^2}{2}\right)$, i.e. we recover the classical log-transform. The resulting distribution of X_t will be precisely normal.
- The process X_t has a constant volatility, so the heavy tails in its *local* distribution are eliminated. However, this desirable property comes at the expense of a potentially more complex trend term: $\hat{\beta}(X_t, t; \boldsymbol{q})$ given by (14) instead of the original $\tilde{\mu}(S_t, t; \boldsymbol{q})$.

Let us use the notation

$$\begin{aligned} X_{i-1}(\boldsymbol{q}) &= f(S_{i-1}; t_{i-1}, \boldsymbol{q}) \ , \\ X_i(\boldsymbol{q}) &= f(S_i; t_i, \boldsymbol{q}) \ , \\ \hat{\beta}_{i-1}(\boldsymbol{q}) &= \tilde{\beta}(S_{i-1}, t_{i-1}, \boldsymbol{q}) = \hat{\beta}(X_{i-1}(\boldsymbol{q}), t_{i-1}, \boldsymbol{q}) \ . \end{aligned}$$

For each time interval (t_{i-1}, t_i), consider the transition density for the process X_t conditioned on the initial value $X_{t_{i-1}} = X_{i-1}(\boldsymbol{q})$:

$$\hat{p}_{i-1}(x, t; \boldsymbol{q}) = \hat{p}(x, t; \boldsymbol{q} \,|\, X_{i-1}(\boldsymbol{q})) \ .$$

Let us assume for the moment that the densities $\hat{p}_{i-1}(x, t; \boldsymbol{q})$ are known.

3.2.2 Likelihood Function for the Normalized Process

It is important to note that one should *not* use the latter densities directly in construction of a maximum likelihood estimator for the parameter vector $\boldsymbol{q}$. I.e., it is *not* possible to simply replace $p_{i-1}(S_i, t_i; \boldsymbol{q})$ with $\hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})$ in (5) or (6) and maximize the resulting likelihood density function $\prod_{i=1}^{N} \hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})$ (or its log) over $\boldsymbol{q}$.

The reason for the above restriction is that, although the processes S_t and X_t are connected via a monotonic transform (12), they are not really symmetrical:

- S_t is a directly observable process whereas X_t is not;
- X_t is connected to S_t via a transform which depends on yet-unknown parameters $\boldsymbol{q}$;
- therefore, in construction of our likelihood functions, we can set all $\{dS_i\}$ to an arbitrary small *constant* value (as we did in Section 1), but we cannot do that for the differentials $\{dX_i\}$.

Therefore, we must still use the original likelihood functions (5) and (6) expressed in terms of the observable quantities $\{S_i\}$. We thus re-construct the original transition densities from the tail-normalized ones.

To that end, consider the likelihood dP_i given by (4) again. We have, due to monotonicity of f in s,

$$\begin{aligned} dP_i(\boldsymbol{q}) &= \text{Prob}\,\{S_i \le S_{t_i} \le S_i + dS_i \,|\, S_{i-1}\} \\ &= p_{i-1}(S_i, t_i; \boldsymbol{q})\, dS_i \\ &= \text{Prob}\,\{f(S_i; t_i, \boldsymbol{q}) \le (S_{t_i}; t_i, \boldsymbol{q}) \le f(S_i + dS_i; t_i, \boldsymbol{q})\} \end{aligned}$$

$$
\begin{aligned}
&= \text{Prob}\left\{ X_i(\boldsymbol{q}) \leq X_{t_i} \leq X_i(\boldsymbol{q}) + \frac{\partial f}{\partial s}(S_i; t_i, \boldsymbol{q})\, dS_i \right\} \\
&= \hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})\, \frac{\partial f}{\partial s}(S_i; t_i, \boldsymbol{q})\, dS_i \\
&= \hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})\, dX_i \ ,
\end{aligned}
$$

where

$$
dX_i \ = \ \frac{\partial f}{\partial s}(S_i; t_i, \boldsymbol{q})\, dS_i \ = \ \frac{dS_i}{\tilde{\sigma}(S_i, t_i; \boldsymbol{q})} \ . \tag{15}
$$

Thus,

$$
p_{i-1}(S_i, t_i; \boldsymbol{q}) = \ \frac{\hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})}{\tilde{\sigma}(S_i, t_i; \boldsymbol{q})} \ , \tag{16}
$$

and this expression can be substituted into (5) and (6). Alternatively, one can write down the likelihood function for the process X_t:

$$
\prod_{i=1}^{N} \hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})\, dX_i \ ,
$$

but, instead of making $\{dX_i\}$ constant, re-write them using (15).

In both cases, the resulting log-likelihood maximization problem will be:

$$
\begin{aligned}
L_1^X(\boldsymbol{q}) \ &= \ \sum_{i=1}^{N} \log\left(\frac{\hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q})}{\tilde{\sigma}(S_i, t_i; \boldsymbol{q})} \right) \ , \\
\boldsymbol{q}^* \ &= \ \arg\max_{\boldsymbol{q} \in \mathcal{Q}} L_1^X(\boldsymbol{q}) \ .
\end{aligned} \tag{17}
$$

3.2.3 Construction of Normalized Transition PDFs

Therefore, we are left with the problem of constructing the transition densities

$$
\hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q}) \ .
$$

The simplest way of doing that is to apply a locally-normal approximation similar to (10) and (11), but now making use of the unitary volatility of the process X_t. This yields

$$
\hat{p}_{i-1}(X_i(\boldsymbol{q}), t_i; \boldsymbol{q}) \ \simeq \ \frac{1}{\sqrt{2\pi\tau_i}} \exp\left\{ -\frac{1}{2\tau_i} \left(X_i(\boldsymbol{q}) - X_{i-1}(\boldsymbol{q}) - \hat{\beta}_{i-1}(\boldsymbol{q})\, \tau_i \right)^2 \right\} \tag{18}
$$

and an estimator

$$
\begin{aligned}
L_1^{\text{LNX}}(\boldsymbol{q}) \ &= \ \sum_{i=1}^{N} \left[\log \tilde{\sigma}(S_i, t_i, \boldsymbol{q}) + \frac{\left(X_i(\boldsymbol{q}) - X_{i-1}(\boldsymbol{q}) - \hat{\beta}_{i-1}(\boldsymbol{q})\, \tau_i \right)^2}{2\tau_i} \right] \ , \\
\boldsymbol{q}^* \ &= \ \arg\min_{\boldsymbol{q} \in \mathcal{Q}} L_1^{\text{LNX}}(\boldsymbol{q}) \ .
\end{aligned}
$$

However, this is still not ideal. What we really want is to take into account the exact term structure of the new trend $\hat{\beta}(s, t; \boldsymbol{q})$, rather than just set it to a fixed initial value within each interval (t_{i-1}, t_i). Several solutions have been proposed to that end, e.g. Hermite polynomial expansions of the transition density [1]. In Section 3.3, we will elaborate a state-of-the-art approximation method for the densities $\hat{p}_{i-1}(x, t; \boldsymbol{q})$, based on the **Heat Kernel** theory.

3.3 Heat Kernels

Let us consider SDE (14) once again, in an interval $t_{i-1} \le t < t_i$. As before, let $\hat{p}_{i-1}(x,t;\boldsymbol{q})$ be the PDF of X_t conditioned on the initial value $X_{t_{i-1}} = X_{i-1}(\boldsymbol{q})$. In general, this function is a solution to the **Fokker-Planck** partial differential equation (PDE) [14]. For SDE (14) with unitary volatility, the Fokker-Planck PDE will be

$$\frac{\partial \hat{p}_{i-1}}{\partial t} = -\frac{\partial(\hat{\beta}(x,t;\boldsymbol{q})\,\hat{p}_{i-1})}{\partial x} + \frac{1}{2}\frac{\partial^2 \hat{p}_{i-1}}{\partial x^2} \ , \tag{19}$$

with the initial condition

$$\hat{p}(x, t_{i-1}; \boldsymbol{q}) = \delta(x - X_{i-1}(\boldsymbol{q})) \ , \tag{20}$$

where $\delta(\cdot)$ is the Dirac delta function, and with the boundary conditions

$$\hat{p}(x = \pm\infty, t; \boldsymbol{q}) = 0 \ . \tag{21}$$

The method of **Heat Kernels** allows us to construct asymptotic approximations of the solutions to the Fokker-Planck PDE (with the above initial and boundary conditions), for small values of $\tau = t - t_{i-1}$. The full theory behind this method is quite complex and is not discussed in this paper; see [2] for details. Here we present the final result for $\hat{p}_{i-1}(x,t;\boldsymbol{q})$ expressed via a Heat Kernel:

$$\begin{aligned} \hat{p}_{i-1}(x,t;\boldsymbol{q}) &= \frac{1}{\sqrt{2\pi\tau}} \exp\left\{ -\frac{(x - X_{i-1}(\boldsymbol{q}))^2}{2\tau} + \int_{X_{i-1}(\boldsymbol{q})}^{x} \hat{\beta}(X, t_{i-1}; \boldsymbol{q})\, dX - \tau\, a_{i-1}^{(1)}(x;\boldsymbol{q}) \right\} \\ &+ O(\tau^{3/2}) \ . \end{aligned} \tag{22}$$

Here the term $a_{i-1}^{(1)}(x;\boldsymbol{q})$ is called the 1st Hadamard-Minakshisundaram-De Witt-Seeley coefficient (*HMDS* for short). It is given by

$$a_{i-1}^{(1)}(x;\boldsymbol{q}) = \frac{\hat{\beta}(x, t_{i-1}; \boldsymbol{q}) - \hat{\beta}_{i-1}(\boldsymbol{q}) + \int_{X_{i-1}(\boldsymbol{q})}^{x} \hat{\beta}^2(X, t_{i-1}; \boldsymbol{q})\, dX}{2(x - X_{i-1}(\boldsymbol{q}))} \tag{23}$$

$$\simeq \frac{1}{2}\left(\frac{\partial \hat{\beta}}{\partial x}(X_{i-1}(\boldsymbol{q}), t_{i-1}; \boldsymbol{q}) + \hat{\beta}_{i-1}^2(\boldsymbol{q}) \right) \quad \text{as} \quad x \to X_{i-1}(\boldsymbol{q}) \ . \tag{24}$$

Remarks:

- It is also possible to consider the Fokker-Planck PDE and the Heat Kernel expansion for the original problem (1). However, in case of non-constant volatility, Heat Kernel construction becomes much more involved. Thus, it is really beneficial to apply the Tail Normalization transform (12) and to work with the stochastic process X_t rather than the original process S_t.

- Although the trend term $\hat{\beta}$ in general depends on t, the asymptotic expansion (22) with an error of order $O(\tau^{3/2})$ has been obtained by evaluating $\hat{\beta}$ at $t = t_{i-1}$ only.

- On the other hand, explicit dependency of $\hat{\beta}$ on x shows up prominently in that expansion, via the following two terms.

  Firstly, the expansion contains the factor

$$\mathcal{P}_{i-1}(x;\boldsymbol{q}) = \exp\left\{ \int_{X_{i-1}(\boldsymbol{q})}^{x} \hat{\beta}(X, t_{i-1}; \boldsymbol{q})\, dX \right\} \tag{25}$$

which is called the *Parallel transport*, and the integral under the exponent is called the *Abelian connection* and denoted by $\mathcal{A}_{i-1}(x;\boldsymbol{q})$. The Parallel transport factor captures the effects of non-trivial dependency of $\hat{\beta}$ on x *"in the large"*, i.e. when x is sufficiently far away from $X_{i-1}(\boldsymbol{q})$.

Secondly, the HMDS coefficient (24) captures the effects of non-trivial dependency of $\hat{\beta}$ on x *"in the small"*, via the derivative $\frac{\partial\hat{\beta}}{\partial x}$.

- The latter coefficient can be written either in its full form (23) derived in [12], or in its approximate ("diagonal") form (24) which is widely known in the literature (see e.g. [2]). Although the full form is more accurate when the points x and $X_{i-1}(\boldsymbol{q})$ are far apart, we will in the following use the *diagonal form*, for the following reasons:
  - the whole Heat Kernel expansion (22) is an approximation only, so the more complex full form may not offer any significant advantages;
  - the diagonal form will be shown to be applicable to Heat Kernels with finite boundaries as well (see Section 4.2.2), and will stay invariant as the boundaries tend to infinity again.
- Therefore, the advent of the Parallel transport and the (1st) HMDS coefficient can be viewed as a principal advantage of using the Heat Kernel method in construction of maximum likelihood estimators, as they allow us to abandon any low-precision (constant-valued) approximations not only in the volatility term (where they are eliminated by means of the Tail Normalization transform), but also in the trend term.
- Also note that if $\hat{\beta} = \hat{\beta}(\boldsymbol{q}) = \text{const}$, we recover a normal density from (22). It becomes

$$\begin{aligned}\hat{p}_{i-1}(x,t;\boldsymbol{q}) &= \frac{1}{\sqrt{2\pi\tau}}\exp\left\{-\frac{(x-X_{i-1}(\boldsymbol{q}))^2}{2\tau}+\hat{\beta}(\boldsymbol{q})\,(x-X_{i-1}(\boldsymbol{q}))-\tau\,\frac{\hat{\beta}^2(\boldsymbol{q})}{2}\right\}\\ &= \frac{1}{\sqrt{2\pi\tau}}\exp\left\{-\frac{1}{2\tau}\left(x-X_{i-1}(\boldsymbol{q})-\hat{\beta}(\boldsymbol{q})\,\tau\right)^2\right\}\ . \qquad (26)\end{aligned}$$

  The term $\int_{X_{i-1}(\boldsymbol{q})}^{x}\hat{\beta}(X,t_{i-1};\boldsymbol{q})\,dX$ becomes $\hat{\beta}(\boldsymbol{q})\,(x-X_{i-1}(\boldsymbol{q}))$ and the term $a_{i-1}^{(1)}(x;\boldsymbol{q})$ becomes $\frac{\hat{\beta}^2(\boldsymbol{q})}{2}$. By analogy with this simple reduction, in Section 4.2.2 we will apply the *inverse* transform from a constant to a non-constant trend $\hat{\beta}$.

■

Finally, one can plug expression (22) into (17) to obtain the actual Heat Kernel-based estimator for the parameter vector $\boldsymbol{q}$.

4 Computation of 3-Variate Transition Densities

Based on the results of Section 3, we will now develop a method for computing 3-variate transition densities (8), to be used in the estimator (9). This method is the main objective of our report.

4.1 Tail Normalization Transform in the 3-Variate Case

We will begin with the Tail Normalization transform (12) and consider the normalized stochastic process X_t given by (13) and satisfying SDE (14). Similar to the 3-dimensional stochastic process

(S_t, L_t, H_t) (see 7), we introduce a 3-dimensional process (X_t, A_t, B_t), where A_t and B_t are the "low" and "high" (resp.) values of X_t since some starting time t_*:

$$A_t = \min_{t_* \le t' \le t} X_t \ , \quad B_t = \max_{t_* \le t' \le t} X_t \ . \tag{27}$$

Remark:
Recall that $X_t = f(S_t; t, \boldsymbol{q})$ where f is given by (12) and is monotonic in its first argument (s). However, since this transform may also depends on t in some arbitrary way, the equalities $A_t = f(L_t; t, \boldsymbol{q})$ or $B_t = f(H_t; t, \boldsymbol{q})$ in general do *not* hold. We will need to pay close attention to this in Section 4.3.
■

The process (X_t, A_t, B_t) is again a Markovian one. Thus, for each time interval (t_{i-1}, t_i) its joint distribution conditioned on the initial triple $(X_{i-1}, X_{i-1}, X_{i-1})$ at the starting time $t_* = t_{i-1}$, is independent from similar distributions for other intervals and is fully determined by X_{i-1}, t_{i-1} and the parameter vector $\boldsymbol{q}$.

For any interval (t_{i-1}, t_i), let us introduce the function $P_{i-1}(x, a, b, t; \boldsymbol{q})$ such that

$$P_{i-1}(x, a, b, t; \boldsymbol{q})\, dx \ = \ \text{Prob}\,\{X_t \in (x, x+dx) \ \&\ a \le A_t \ \&\ B_t \le b \ \mid X_{i-1}\} \ . \tag{28}$$

Thus, $P_{i-1}(x, a, b, t; \boldsymbol{q})\, dx$ is the probability of X_t falling into the infinitesimal interval $(x, x+dx)$, while for all $t_{i-1} \le t' \le t$, $X_{t'}$ stays in the range (a, b), conditional on $X_{t_{i-1}} = X_{i-1}$.

By itself, $P_{i-1}(x, a, b, t; \boldsymbol{q})\, dx$ is not a transition PDF. However, if $\hat{p}_{i-1}(x, a, b, t; \boldsymbol{q})$ is the corresponding 3-variate transition PDF (also conditional on $X_{t_{i-1}} = X_{i-1}$), then it follows from the definition of the multi-variate PDF that the probability $P_{i-1}(x, a, b, t; \boldsymbol{q})\, dx$ can be expressed as

$$\text{Prob}\,\{X_t \in (x, x+dx) \ \&\ A_t \ge a \ \&\ B_t \le b \ \mid X_{i-1}\} \ = \ \int_a^{+\infty} dA \int_{-\infty}^{b} dB\ \hat{p}_{i-1}(x, A, B, t; \boldsymbol{q})\, dx \ ,$$

and therefore

$$\hat{p}_{i-1}(x, a, b, t; \boldsymbol{q}) = -\frac{\partial^2 P_{i-1}(x, a, b, t; \boldsymbol{q})}{\partial a\, \partial b} \ . \tag{29}$$

Obviously, if $a > b$, we must have $\hat{p}_{i-1}(x, a, b, t; \boldsymbol{q}) = P_{i-1}(x, a, b, t; \boldsymbol{q}) = 0$. If $a = -\infty$ and $b = +\infty$, then $\hat{p}_{i-1}(x, a, b, t; \boldsymbol{q})$ coincides with the 1-variate transition PDF $\hat{p}(x, t; \boldsymbol{q})$ which can be approximated by the Heat Kernel (22).

4.2 Fokker-Planck PDE in the 3-Variate Case

We now consider the problem of computing the function $P_{i-1}(x, a, b, t; \boldsymbol{q})$. According to [7], this function satisfies the same Fokker-Planck PDE (19) in x and t (just replace $\hat{p}_{i-1}$ with P_{i-1}), with the same initial condition (20), and with new "finite" boundary conditions

$$P_{i-1}(x = a, a, b, t; \boldsymbol{q}) \ = \ P_{i-1}(x = b, a, b, t; \boldsymbol{q}) \ = \ 0 \tag{30}$$

instead of "infinite" boundary conditions (21).

If the trend $\hat{\beta}$ given by (14) is constant, this PDE problem can be solved analytically. There are at least two methods of doing that [7], which are discussed below.

4.2.1 Method 1: Separation of Variables: Sturm-Liouville Problem

This is a classical method. We will seek a solution to PDE (19) in the form of a separable function of x and t:

$$P_{i-1}(x, a, b, t; \boldsymbol{q}) = P_1(x)\, P_2(t) \ ;$$

here both P_1 and P_2 will also depend on a, b and $\boldsymbol{q}$, but we will temporarily omit those dependencies for brevity (also in $\hat{\beta}(x)$). By substituting the above expression into equation (19) and boundary conditions (30), we get

$$\frac{\dot{P}_2}{P_2} = \frac{\frac{1}{2}P_1'' - (\hat{\beta}P_1)'}{P_1} \ ,$$

where the dot ($\dot{}$) denotes differentiation w.r.t. t and the prime ($'$) denotes differentiation w.r.t x. If $\hat{\beta}$ does not depend on t, then the right-hand side (RHS) of the above equality is a function of only x while the left-hand-side (LHS) is a function of only t. Therefore, both sides must be equal to a (yet unknown) constant which will be denoted by $\left(-\frac{\lambda}{2}\right)$.

We then get the following ordinary differential equations (ODEs) instead of the original PDE:

$$\begin{aligned} &\frac{1}{2}P_1'' - (\hat{\beta}P_1)' + \frac{\lambda}{2}P_1 = 0 \ , \\ &\dot{P}_2 - \frac{\lambda}{2}P_2 = 0 \ . \end{aligned}$$

The ODE for $P_1(x)$ is to be considered together with the boundary conditions

$$P_1(a) = P_1(b) = 0 \ ;$$

we are looking for non-trivial solutions ($P1 \not\equiv 0$). Note that $a \le X_{i-1} \le b$ and $a \le x \le b$. Let us make a substitution

$$P_1(x) = \exp\left(\int_{X_{i-1}}^{x} \hat{\beta}(X)\, dX\right) y(x) \ ,$$

where the exponential factor in $P_1(x)$ is nothing else but the Parallel transport (25). Then

$$\begin{aligned} &y'' + \Big(\lambda - (\hat{\beta}^2(x) + \hat{\beta}'(x))\Big)\, y = 0 \ , \\ &y(a) = y(b) = 0 \ . \end{aligned} \tag{31}$$

This is a famous **Sturm-Liouville problem** [10]. For this problem, it is known that under some genericity conditions, there exists an infinite sequence of discrete *eigenvalues* λ_n, where $n \in \mathbb{N}$, and the corresponding *eigenfunctions* $y_n(x) \not\equiv 0$ satisfying the boundary-value problem (31) with $\lambda = \lambda_n$. The eigenvalues and eigenfunctions actually depend on the function $\hat{\beta}$, boundaries a and b and the model parameters $\boldsymbol{q}$. Obviously, eigenfunctions $y_n(x)$ are defined up to constant coefficients; we will assume that they are normalized: $||y_n||_{L_2(a,b)} = 1$. As $n \to +\infty$, the following asymptotic relationships hold:

$$\begin{aligned} \lambda_n &\simeq \left(\frac{\pi n}{b-a}\right)^2 + 2\,\alpha + O\left(\frac{1}{n}\right) , \\ y_n(x) &\simeq \sqrt{\frac{2}{b-a}} \sin\left(\frac{\pi n(x-a)}{b-a}\right) + O\left(\frac{1}{n}\right) \ , \end{aligned} \tag{32}$$

where the term α is very similar to the 1st HMDS coefficient (24):

$$\alpha \;=\; \frac{\hat{\beta}(b) - \hat{\beta}(a) + \int_a^b \hat{\beta}^2(X)\,dX}{2(b-a)} \;\simeq\; \frac{1}{2}\left(\hat{\beta}'(X_{i-1}) + \hat{\beta}^2(X_{i-1})\right) \;. \tag{33}$$

The similarity between expressions (23), (24) and (33) is of course no coincidence: both expressions originate from transformation of the Fokker-Planck differential operator into a canonical form which does not contain the 1st derivative w.r.t x. Note that the second (local) form of α does not depend on the boundaries a and b, similar to the diagonal form of the 1st HMDS coefficient which is also evaluated locally at X_{i-1}. Thus, both the Parallel transport and the 1st HMDS coefficient which appeared in our Heat Kernel construction, are also used in the Separation of Variables method.

Then for each eigenvalue λ_n and the corresponding eigenfuction $y_n(x)$, we get the following solution for $P_1(x)$ and $P_2(t)$ (up to constant coefficients):

$$\begin{aligned} P_{1n}(x) &= \exp\left(\int_{X_{i-1}}^{x} \hat{\beta}(X)\,dX\right) y_n(x) \;, \\ P_{2n}(t) &= \exp\left(-\frac{\lambda_n \tau}{2}\right) \;. \end{aligned}$$

Here $\tau = t - t_{i-1}$ as usual. The general solution for $P_{i-1}(x, a, b, t; \boldsymbol{q})$ can then be written as *Fourier series* over those eigenfunctions:

$$P_{i-1}(x, a, b, t; \boldsymbol{q}) \;=\; \sum_{n-1}^{+\infty} C_n\, P_{1n}(x)\, P_{2n}(t) \;=\; \exp\left(\int_{X_{i-1}}^{x} \hat{\beta}(X)\,dX\right) \sum_{n=1}^{+\infty} C_n \exp\left(-\frac{\lambda_n \tau}{2}\right) y_n(x) \;,$$

where $\{C_n\}$ are some constant coefficients. It remains to determine the latter coefficients to satisfy the initial condition (20). By setting $\tau = 0$ in the above equation, we get

$$\sum_{n=1}^{+\infty} C_n\, y_n(x) \;=\; \exp\left(-\int_{X_{i-1}}^{x} \hat{\beta}(X)\,dX\right) \delta(x - X_{i-1}) \;.$$

Since the eigenfunctions $\{y_n(x)\}$ constitute an orthogonal normalized system of functions in the interval (a, b), we obtain the expressions

$$C_n = \int_a^b \exp\left(-\int_{X_{i-1}}^{x} \hat{\beta}(X)\,dX\right) y_n(x)\, \delta(x - X_{i-1})\,dx \;=\; y_n(X_{i-1}) \;,$$

due to the well-known property of the Dirac delta function. Therefore, we get the expression

$$P_{i-1}(x, a, b, t; \boldsymbol{q}) \;=\; \exp\left(\int_{X_{i-1}}^{x} \hat{\beta}(X; \boldsymbol{q})\,dX\right) \sum_{n=1}^{+\infty} \exp\left(-\frac{\lambda_n\, \tau}{2}\right) y_n(X_{i-1}(\boldsymbol{q}))\, y_n(x) \;, \tag{34}$$

where $\{\lambda_n\}$ and $\{y_n(x)\}$ $(n \in \mathbb{N})$ are the eigenvalues and the corresponding normalized eigenfunctions of the Sturm-Liouville problem (31); they depend on $\hat{\beta}(x; \boldsymbol{q})$, a and b, and they are asymptotically distributed according to (32).

Unfortunately, in general, the eigenvalues and eigenfunctions are very rarely available in a closed analytical form. Though various numerical methods for computing them are available, they may be too computationally-expensive for the model calibration purposes, as the eigenvalues and eigenfunctions will eventually depend on the variable vector of parameters $\boldsymbol{q}$.

For this reason, we take an alternative approach: we will approximate $\{\lambda_n\}$ and $\{y_n(x)\}$ by the main terms of the asymptotic expansions (32). Furthermore, although the Separation of Variables method requires that $\hat{\beta}$ must not depend on t, we will allow such dependencies as well: By analogy with the Heat Kernel expansion (22), we will fix $t = t_{i-1}$ is $\hat{\beta}$ (but obviously not $x = X_{i-1}$). We finally get the following approximation (using the local form of α):

$$\begin{aligned} P_{i-1}(x,a,b,t;\boldsymbol{q}) \;\simeq\; & \frac{2}{b-a}\exp\left\{\int_{X_{i-1}}^{x}\hat{\beta}(X,t_{i-1};\boldsymbol{q})\,dX\right. \\ & \left.-\frac{t}{2}\left(\frac{\partial\hat{\beta}}{\partial x}(X_{i-1}(\boldsymbol{q}),t_{i-1};\boldsymbol{q})+\hat{\beta}^2(X_{i-1}(\boldsymbol{q}),t_{i-1};\boldsymbol{q})\right)\right\} \\ \times\; & \sum_{n=1}^{+\infty}\exp\left(-\frac{\tau}{2}\left(\frac{\pi n}{b-a}\right)^2\right)\sin\left(\frac{\pi n(X_{i-1}(\boldsymbol{q})-a)}{b-a}\right)\sin\left(\frac{\pi n(x-a)}{b-a}\right)\;, \\ \text{where}\quad \tau \;=\; & t-t_{i-1}\;. \end{aligned} \tag{35}$$

Remarks:

- If $\hat{\beta} = \hat{\beta}(\boldsymbol{q}) = \text{const}$, then the eigenvalues and eigenfunctions of (31) can easily be obtained analytically, and they will coincide *exactly* with the approximations used in (35). We thus recover the result for a "Brownian motion with constant drift" presented in [7].

- In general, it is possible to use higher-order approximations for eigenvalues and eigenfunctions, instead of (32). E.g., [8] provides formulas for unnormalized eigenfunctions for the Sturm-Liouville problem (31) with an error of order $O\left(\frac{1}{n^2}\right)$. However, in our case the pros and cons of using such higher-order approximations are uncertain; in particular, they would result in much more complex expressions for *normalized* eigenfunctions. For this reason, we currently use approximations (32) only.

- Furthermore, asymptotic error orders in (32) or similar formulas in [8] do not tell us anything about the accuracy of eigenvalues and eigenfunctions approximation for *small* values of n; whereas small-n terms provide major contribution to the sum of series in (35). Nevertheless, we consider the approximation (35) to be adequate because it correctly recovers the known result for $\hat{\beta} = \text{const}$.

- However, the main problem with approximation (35) is slow convergence of series for small values of τ (whereas for large values of τ, fixing $t = t_{i-1}$ in $\hat{\beta}$ would no longer be valid). Indeed, the RHS of (35) becomes the Dirac delta function as $\tau \to 0$.

  In the Heat Kernel (22), the RHS also tends to the delta function as $\tau \to 0$, but that property is captured via the multiplicative term

  $$\frac{1}{\sqrt{2\pi\tau}}\exp\left\{-\frac{(x-X_{i-1}(\boldsymbol{q}))^2}{2\tau}\right\}$$

  rather than via Fourier series. Therefore, it would be advantageous to construct an approximation for $P_{i-1}(x,a,b,t;\boldsymbol{q})$ is the form of a Heat Kernel. This is done in Section 4.2.2 below.

- Further research is required in order to obtain rigorous asymptotic estimates, as functions of τ, for errors in (35) originating from fixing $t = t_{i-1}$ in $\hat{\beta}$ and using low-order approximations (32) for eigenvalues and eigenfunctions. Rough estimates show that the error may be of order $O(\tau)$, as opposed to $O(\tau^{3/2})$ in (22), mainly due to approximation of eigenfunctions.

■

4.2.2 Method 2: Heat Kernels with Finite Boundary Conditions

The Heat Kernel (22) was constructed as a solution to the Fokker-Planck PDE with "infinite" boundary conditions, that is, for the domain $x \in \mathbb{R}$ and the solution vanishing at $x = \pm\infty$. In this section, we will construct a similar Heat Kernel-based solution for the domain $a \le x \le b$ and the solution vanishing at finite boundaries $x = a,\, b$.

Technically, this can be done in a number of ways, all of them providing various degrees of approximation. First of all, for $\hat\beta = \text{const}$, it follows from the formula given in [7] that

$$\begin{aligned} P_{i-1}(x,a,b,t;\boldsymbol{q}) &= \frac{1}{\sqrt{2\pi\tau}} \exp\left\{\hat\beta(\boldsymbol{q})(x - X_{i-1}(\boldsymbol{q})) - \tau\,\frac{\hat\beta^2(\boldsymbol{q})}{2}\right\} \\ &\times \sum_{n=-\infty}^{+\infty}\left\{\exp\left[-\frac{(x - X_{i-1}(\boldsymbol{q}) - 2n(b-a))^2}{2\tau}\right]\right. \\ &\left. - \exp\left[-\frac{(x + X_{i-1}(\boldsymbol{q}) - 2n(b-a) - 2a)^2}{2\tau}\right]\right\}\,. \end{aligned} \tag{36}$$

The result of [7] which we used here was constructed using the *"Reflection principle"* for the Brownian motion [14], or equivalently, the *"Method of images"* for solving the Fokker-Planck PDE [16]. Indeed, the terms $2n(b-a)$ in (36) correspond to paths from $X_{i-1}(\boldsymbol{q})$ to x containing multiple reflections from the boundary points a and b. Importantly, because of those terms, the series in (36) are expected to converge rapidly. Also note that the original result of [7] was constructed for the case $X_{i-1} = 0$; formula (36) was obtained from it by translation $x \mapsto x - X_{i-1}$, $a \mapsto a - X_{i-1}$, $b \mapsto b - X_{i-1}$.

We will now extend formula (36) to the general case of $\hat\beta(x,t;\boldsymbol{q})$ using analogies developed in Sections 3.3 and 4.2.1. First, we again fix $t = t_{i-1}$ in $\hat\beta$. Then we re-write $\hat\beta(\boldsymbol{q})(x - X_{i-1}(\boldsymbol{q}))$ as the Abelian connection and $\frac{\hat\beta^2(\boldsymbol{q})}{2}$ as local-form α given by (33). We get:

$$\begin{aligned} P_{i-1}(x,a,b,t;\boldsymbol{q}) &\simeq \frac{1}{\sqrt{2\pi\tau}} \exp\left\{\int_{X_{i-1}(\boldsymbol{q})}^{x} \hat\beta(X,t_{i-1};\boldsymbol{q})\,dX - \tau\,\alpha_{i-1}(\boldsymbol{q})\right\} \\ &\times \sum_{n=-\infty}^{+\infty}\left\{\exp\left[-\frac{(x - X_{i-1}(\boldsymbol{q}) - 2n(b-a))^2}{2\tau}\right]\right. \\ &\left. - \exp\left[-\frac{(x - X_{i-1}(\boldsymbol{q}) - 2n(b-a) - 2a)^2}{2\tau}\right]\right\}\,, \end{aligned} \tag{37}$$

$$\begin{aligned} \text{where}\quad \alpha_{i-1}(\boldsymbol{q}) &= \frac{1}{2}\left(\frac{\partial\hat\beta}{\partial x}(X_{i-1}(\boldsymbol{q}),t_{i-1};\boldsymbol{q}) + \hat\beta^2(X_{i-1}(\boldsymbol{q}),t_{i-1};\boldsymbol{q})\right)\,, \\ \tau &= t - t_{i-1}\,. \end{aligned}$$

Note that if we now let $a \to -\infty$ and $b \to +\infty$ in (37), then the only term in the sum over n which does not vanish is $n = 0$, and we recover the 1-variate Heat Kernel expression (22). This would not be guaranteed, however, if the non-local form of α (similar to the full form of the 1st HMDS coefficient) were used, because it may explicitly depend on a and b, and its asymptotic behavior as $a \to -\infty$ and $b \to +\infty$ is completely arbitrary.

On the other hand, a similar approximation can be constructed using the Heat Kernel methods presented in [2, Section 5.11.5], which are again based on the Reflection principle. The resulting Heat Kernel contains the terms corresponding to $n = 0$ and $n = 1$, and has an asymptotic error order $O(\tau^{3/2})$ similar to (22). From that we can conclude that formulas (37) also provide an error order of at least $o(\tau)$.

4.3 The 3-Variate Calibration Algorithm

We can now assume that the function $P_{i-1}(x,a,b,t;\boldsymbol{q})$ is available, either via approximation (35) or via (37). Let us complete the construction of our 3-variate MLE method.

First of all, we obtain a 3-variate PDF $\hat{p}_{i-1}(x,a,b,t;\boldsymbol{q})$ for the stochastic process (X_t, A_t, B_t) introduced by (27), from $P_{i-1}(x,a,b,t;\boldsymbol{q})$ using (29). For the same reasons as discussed in Section 3.2, we cannot use that PDF $\hat{p}_{i-1}(x,a,b,t;\boldsymbol{q})$ by itself in construction of our likelihood function; we need to re-construct the original transition likelihood

$$p_{i-1}(s,l,h,t;\boldsymbol{q})\,ds\,dl\,dh$$

from the tail-normalized transition likelihood

$$\hat{p}_{i-1}(x,a,b,t;\boldsymbol{q})\,dx\,da\,db$$

and use the former in the optimization procedure (9).

To that end, recall from Section 4.1 that $X_t = f(S_t;t,\boldsymbol{q})$, where f is given by (12), and therefore

$$dx = \frac{\partial f}{\partial s}\,ds = \frac{ds}{\tilde{\sigma}(s,t;\boldsymbol{q})}$$

which is similar to (15). Yet in general, there is no explicit functional dependency connecting L_t and A_t (and similarly, H_t and B_t), due to explicit dependency of the transform f on t.

However, recall from (12) that f depends on t only via the volatility $\tilde{\sigma}$. In typical stochastic models found in Quantitative Finance, either $\tilde{\sigma}$ does not explicitly depend on t at all, or its dependency on t is very weak (e.g. in HJM-type interest rate models [3]), so that within each particular interval (t_{i-1},t_i), we can regard $\tilde{\sigma}$ and f to be constant functions of the explicit argument t.

Thus, let us **assume** that the following differential relationships hold:

$$\begin{aligned}
dx &= \frac{\partial f}{\partial s}(s;t_i,\boldsymbol{q})\,ds = \frac{ds}{\tilde{\sigma}(s;t_i,\boldsymbol{q})}\,,\\
da &= \frac{\partial f}{\partial s}(l;t_i,\boldsymbol{q})\,dl = \frac{dl}{\tilde{\sigma}(l;t_i,\boldsymbol{q})}\,,\\
db &= \frac{\partial f}{\partial s}(h;t_i,\boldsymbol{q})\,dh = \frac{dh}{\tilde{\sigma}(h;t_i,\boldsymbol{q})}\,.
\end{aligned}$$

(It is more convenient for our purposes to fix $t=t_i$ rather than $t=t_{i-1}$ in the above formulas, because we are evaluating the PDFs at $t=t_i$, although the exact value of t is quite irrelevant).

Then, similar to (16), we get the following approximation:

$$p_{i-1}(s,l,h,t;\boldsymbol{q}) \;\simeq\; \frac{\hat{p}_{i-1}(f(s;t_i,\boldsymbol{q}),\,f(l;t_i,\boldsymbol{q}),\,f(h;t_i,\boldsymbol{q}),\,t;\,\boldsymbol{q})}{\tilde{\sigma}(s,t_i;\boldsymbol{q})\,\tilde{\sigma}(l,t_i;\boldsymbol{q})\,\tilde{\sigma}(h,t_i;\boldsymbol{q})}\,,$$

and we finally arrive at the following 3-variate calibration procedure:

$$\begin{aligned}
L_3^X(\boldsymbol{q}) &= \sum_{i=1}^{N}\log\left(\frac{\hat{p}_{i-1}(f(S_i;t_i,\boldsymbol{q}),\,f(L_i;t_i,\boldsymbol{q}),\,f(H_i;t_i,\boldsymbol{q}),\,t_i;\,\boldsymbol{q})}{\tilde{\sigma}(S_i,t_i;\boldsymbol{q})\,\tilde{\sigma}(L_i,t_i;\boldsymbol{q})\,\tilde{\sigma}(H_i,t_i;\boldsymbol{q})}\right)\,,\\
\boldsymbol{q}^* &= \arg\max_{\boldsymbol{q}\in\mathcal{Q}} L_3^X(\boldsymbol{q})\,,
\end{aligned}\tag{38}$$

where the function $\hat{p}_{i-1}$ is given by (29) and the function P_{i-1} is given by either (37), which is a recommended way of computing it, or by (35).

5 Model Validation

After the parameters of model (1) have been calibrated using any of the methods discussed above, we should pose the question of whether the given model with the obtained set of parameters $\boldsymbol{q}$ adequately describes, after all, the stochastic dynamics of the observed process S_t. This is a model validation problem, which is a vast and highly important topic. Without going into any depth here, let us propose a very simple model validation test.

Let us extract the normalized increments Δw_t of a Brownian motion W_t from the time series $\{S_i\}$. By using the locally-normal approximation, we can write

$$\Delta w_i \;=\; \frac{W_{t_i} - W_{t_{i-1}}}{\sqrt{\tau_i}} \;\simeq\; \frac{S_i - S_{i-1} - \tilde{\mu}(S_{i-1}, t_{i-1}; \boldsymbol{q})\,\tau_i}{\tilde{\sigma}(S_{i-1}, t_{i-1}; \boldsymbol{q})\,\sqrt{\tau_i}} \;.$$

Or, if the Tail Normalization transform (12) is used, we similarly get

$$\Delta w_i \;\simeq\; \frac{X_i(\boldsymbol{q}) - X_{i-1}(\boldsymbol{q}) - \hat{\beta}_{i-1}(\boldsymbol{q})\,\tau_i}{\sqrt{\tau_i}} \;.$$

If our stochastic model is valid, the random variables $\{\Delta w_i\}$ must constitute a "residual" Random Walk. In particular, they must all be independent and identically distributed as $\mathcal{N}(0,1)$. Therefore, we can apply a Random Walk test to $\{\Delta w_i\}$ in an attempt to falsify (disprove) the validity of our model.

There are multiple statistical test for the random walk property; the most commonly used family of tests are Variance Ratio (VR) tests [4]. Without providing the details here, we would recommend using the modern powerful multi-period Chen and Deo test originally presented in [5]. The test is based on a Discrete Fourier transform of $\{\Delta w_i\}$.

6 Example: Non-Stationary Ornstein-Uhlenbeck Process

6.1 Transition PDFs

The Ornstein-Uhlenbeck (OU) SDE is a well-known model of mean-reverting stochastic processes, e.g. Interest Rates (where it is called the Vasicek model [14, 3]), Commodity prices or log-returns of Commodity prices [6]. In this section, we consider the Non-Stationary Ornstein-Uhlenbeck process (NSOU). It differs from the classical OU process by a non-constant mean-reversion target, which we will assume to be a linear function of time t. This is a reasonable approximation when the mean-reversion target is a slowly-changing function of t. The SDE of the NSOU process is

$$dS_t = \varkappa(\theta + \mu(t - t_*) - S_t)\,dt + \sigma\,dW_t \;, \tag{39}$$

where the vector of parameters $\boldsymbol{q}$ is 4-dimensional: $\boldsymbol{q} = [\varkappa, \theta, \mu, \sigma^2]$. (Note that by convention, we put σ^2, not σ itself, into the vector $\boldsymbol{q}$). Here $\varkappa > 0$ is the mean-reversion rate, $\theta + \mu(t - t_*)$ is the mean-reversion target, σ^2 is a constant variance (whereas $\sigma > 0$ is a volatility), and t_* is any fixed time instant. The classical (asymptotically stationary) OU process corresponds to the case $\mu = 0$.

SDE (39) can be integrated explicitly with any initial condition $S_{t_{i-1}} = S_{i-1}$. First, let us drop the stochastic term in (39) and consider the linear ODE

$$\frac{dS}{dt} = -\varkappa\,S + \varkappa(\theta + \mu(t - t_*)) \;.$$

It can easily be solved:

$$S_t = C\,\exp(-\varkappa(t - t_{i-1})) + \theta + \mu(t - t_*) - \frac{\mu}{\varkappa} \;,$$

where C is an arbitrary constant. (It is important not to confuse t_{i-1} with t_*: the former corresponds to the initial condition within a given time interval, whereas the latter is a fixed model parameter common for all time intervals).

We now apply the Variation of Constant method: Consider C to be a function of t and W_t rather than a constant, and obtain an SDE for C_t in lieu of (39) by applying the Itô–Döblin differentiation formula. We get

$$dC_t \;=\; \sigma\,\exp(\varkappa(t-t_{i-1}))\,dW_t$$

and thus

$$C_t \;=\; C_{i-1} + \sigma\int_{t_{i-1}}^{t}\exp(\varkappa(u-t_{i-1}))\,dW_u\ ,$$

where C_{i-1} is a constant which depends on the initial condition S_{i-1}. Therefore,

$$S_t \;=\; C_{i-1}\exp(-\varkappa(t-t_{i-1})) + \sigma\int_{t_{i-1}}^{t}\exp(-\varkappa(t-u))\,dW_u + \theta + \mu(t-t_*) - \frac{\mu}{\varkappa}\ .$$

By substituting the initial condition $S_{t_{i-1}} = S_{i-1}$ we get

$$C_{i-1} = S_{i-1} - \mu T_{i-1} + \frac{\mu}{\varkappa} - \theta\ ,$$

where we use the notation

$$T_{i-1} = t_{i-1} - t_*\ ,$$

and finally

$$\begin{aligned} S_t \;&=\; \left(S_{i-1} - \mu T_{i-1} + \frac{\mu}{\varkappa} - \theta\right)\exp(-\varkappa\tau) \\ &+\; \mu(t-t_*) + (1-\exp(-\varkappa\tau))\left(\theta - \frac{\mu}{\varkappa}\right) \\ &+\; \sigma\int_{t_{i-1}}^{t}\exp(-\varkappa(t-u))\,dW_u\ , \end{aligned} \tag{40}$$

where $\tau = t - t_{i-1}$ as before.

It follows from (41) and the properties of the Itô integral [14] that S_t (conditional on S_{i-1}) is **normally distributed** with the mean equal to

$$\left(S_{i-1} - \mu T_{i-1} + \frac{\mu}{\varkappa} - \theta\right)\exp(-\varkappa\tau) + \mu(t-t_*) + (1-\exp(-\varkappa\tau))\left(\theta - \frac{\mu}{\varkappa}\right)$$

and the standard deviation equal to

$$\sigma\sqrt{\int_{t_{i-1}}^{t}\exp(-2\varkappa(t-u))du} \;=\; \sigma\sqrt{\frac{1}{2\varkappa}(1-\exp(-2\varkappa\tau))}\ .$$

Therefore, a price process S_t which follows the NSOU SDE (39) can take arbitrary values, both positive and negative, so formally speaking, it is not suitable for modeling the dynamics of strictly-positive price processes. However, if the probability of S_t taking a negative value within the interval $(t_{i-1}, t_{i-1}+\tau)$ is negligible, then the NSOU process may still be a reasonable modeling assumption even for positive-definite price processes; in any case, it could be a valid model for the dynamics of price log-returns.

For small values of $\varkappa\tau$ we can approximate the normal distribution for S_t by the following one:

$$S_t \;\approx\; \mathcal{N}\left(S_{i-1} + \varkappa(\theta + \mu(t-t_*) - S_{i-1})\,\tau,\ \sigma\sqrt{\tau}\right)\ .$$

This is the same as locally-normal approximation (11).

Because NSOU is a constant-volatility process, the Tail Normalization transform is trivial, and is essentially equivalent to the locally-normal approximation: (12) becomes

$$x = f(s;\boldsymbol{q}) = \frac{s}{\sigma} \ ,$$

that is, the normalized process X_t is obtained from S_t simply by division by σ. The trend of X_t is then

$$\hat{\beta}(x,t;\boldsymbol{q}) = \varkappa\left(\frac{\theta+\mu(t-t_*)}{\sigma} - x\right) \ ,$$

(it can also be obtained from (14)), and we would expect that its 1-variate transition PDF for small values of τ will be

$$\hat{p}_{i-1}(x,t;\boldsymbol{q}) \ \simeq \ \frac{1}{\sqrt{2\pi\tau}}\exp\left\{-\frac{1}{2\tau}\left(x - X_{i-1}(\boldsymbol{q}) - \varkappa\left(\frac{\theta+\mu T_{i-1}}{\sigma} - X_{i-1}(\boldsymbol{q})\right)\right)^2\right\} \ , \qquad (41)$$

where obviously

$$X_{i-1}(\boldsymbol{q}) \ = \ \frac{S_{i-1}}{\sigma} \ .$$

Let is now compute the Abelian connection:

$$\begin{aligned}\mathcal{A}_{i-1}(x;\boldsymbol{q}) \ &= \ \int_{X_{i-1}}^{x}\hat{\beta}(X;\boldsymbol{q})\,dX \ = \ \varkappa(x - X_{i-1}(\boldsymbol{q}))\left(\frac{\theta+\mu T_{i-1}}{\sigma} - \frac{x + X_{i-1}(\boldsymbol{q})}{2}\right)\\ &= \ \frac{\varkappa}{\sigma^2}(\sigma x - S_{i-1})\left(\theta + \mu T_{i-1} - \frac{\sigma x + S_{i-1}}{2}\right) \qquad (42)\end{aligned}$$

and also compute the 1st HMDS coefficient defined in (24); in its diagonal (local) form, it is evaluated at $X_{i-1}(\boldsymbol{q})$ only and does not depend explicitly on x:

$$\begin{aligned}a^{(1)}_{i-1}(\boldsymbol{q}) \ &= \ -\frac{\varkappa}{2}\left(1 - \varkappa\left(\frac{\theta+\mu T_{i-1}}{\sigma} - X_{i-1}(\boldsymbol{q})\right)\right)\\ &= \ -\frac{\varkappa}{2}\left(1 - \frac{\varkappa}{\sigma}(\theta + \mu T_{i-1} - S_{i-1})\right) \ . \qquad (43)\end{aligned}$$

Note that in expressions (41), (42) and (43) we fix $t = t_{i-1}$ in time-dependent right-hand sides, in accordance with Section 3.3; in (41), the only explicit dependency on t is via $\tau = t - t_{i-1}$.

We can then compute the 1-variate transition PDF based on the Heat Kernel expansion (22):

$$\hat{p}_{i-1}(x,t;\boldsymbol{q}) \ \simeq \ \frac{1}{\sqrt{2\pi\tau}}\exp\left\{-\frac{(x - X_{i-1})^2}{2\tau} + \mathcal{A}_{i-1}(x;\boldsymbol{q}) - \tau\, a^{(1)}_{i-1}(\boldsymbol{q})\right\} \ . \qquad (44)$$

Note that although both formulas (41) and (44) describe normal PDFs, they are *not* functionally equivalent. This is in contrast to the case of a Brownian Motion with a constant drift (26) when the actual transition PDF and its Heat Kernel approximation were exactly identical.

Finally, the parameter vector $\boldsymbol{q}^*$ for the NSOU process (39) can be obtained from historical market data using the Heat Kernel-based "Open-Close" (1-variate) calibration procedure (17) or the "OHLC" (3-variate) calibration procedure (38).

6.2 Barrier Crossing Probabilities

In the following, we will also require the probability P^+_{i-1} conditional on the "open" price S_{i-1}, that the asset price will exceed $S_{i-1} + \Delta S$ (a "high barrier") at some time t' within a given time horizon Δt:

$$\begin{aligned} P^+_{i-1} &= \text{Prob}\left\{\exists\, t' \in (t_{i-1},\, t_{i-1}+\Delta t) \,:\, S_{t'} > S_{i-1} + \Delta S\right\} \\ &= \text{Prob}\left\{\max_{t_{i-1}<t'\le t_{i-1}+\Delta t} S_{t'} > S_{i-1} + \Delta S\right\} \\ &= \text{Prob}\left\{\max_{t_{i-1}<t'\le t_{i-1}+\Delta t} X_{t'} > X_{i-1} + \frac{\Delta S}{\sigma}\right\} \,, \end{aligned}$$

where $X_{i-1} = \frac{S_{i-1}}{\bar\sigma}$ as before. Note that Δt does not need to be equal to the interval length τ_i.

Similarly, we define P^-_{i-1} as the probability, conditional on S_{i-1}, that the asset price will fall below $S_{i-1} - \Delta S$ (a "low barrier") within time horizon Δt:

$$\begin{aligned} P^-_{i-1} &= \text{Prob}\left\{\exists\, t' \in (t_{i-1},\, t_{i-1}+\Delta t) \,:\, S_{t'} < S_{i-1} - \Delta S\right\} \\ &= \text{Prob}\left\{\min_{t_{i-1}<t'\le t_{i-1}+\Delta t} S_{t'} < S_{i-1} - \Delta S\right\} \\ &= \text{Prob}\left\{\min_{t_{i-1}<t'\le t_{i-1}+\Delta t} X_{t'} < X_{i-1} - \frac{\Delta S}{\sigma}\right\} \,. \end{aligned}$$

In order to calculate the probabilities $P^\pm_{i-1}$, we can re-use the function $P_{i-1}(x, a, b, t; \boldsymbol{q})$ given by (37). Note that using an alternative formula (35) is not recommended because of slow convergence. From definition (28) it is easy to see that

$$\begin{aligned} P^+_{i-1} &= 1 - \int_{-\infty}^{\frac{S_{i-1}+\Delta S}{\sigma}} P_{i-1}\left(x,\, -\infty,\, \frac{S_{i-1}+\Delta S}{\sigma},\, t_{i-1}+\Delta t;\, \boldsymbol{q}\right) dx \\ &= 1 - \frac{1}{\sigma}\int_{-\infty}^{S_{i-1}+\Delta S} P_{i-1}\left(\frac{s}{\sigma},\, -\infty,\, S_{i-1}+\Delta S,\, t_{i-1}+\Delta t;\, \boldsymbol{q}\right) ds \,, \\ P^-_{i-1} &= 1 - \int_{\frac{S_{i-1}-\Delta S}{\sigma}}^{+\infty} P_{i-1}\left(x,\, \frac{S_{i-1}-\Delta S}{\sigma},\, +\infty,\, t_{i-1}+\Delta t;\, \boldsymbol{q}\right) dx \\ &= 1 - \frac{1}{\sigma}\int_{S_{i-1}-\Delta S}^{+\infty} P_{i-1}\left(\frac{s}{\sigma},\, S_{i-1}-\Delta S,\, +\infty,\, t_{i-1}+\Delta t;\, \boldsymbol{q}\right) ds \,. \end{aligned}$$

Thus, $a = -\infty$ in P^+_{i-1} and $b = +\infty$ in P^-_{i-1}. In order to evaluate these respective probabilities, we retain only those terms in (37) which do not depend on a (resp. b), because all other terms will vanish as $a \to -\infty$ (resp. $b \to +\infty$). Finally, by substituting the Abelian connection (42) and the 1st HMDS coefficient (43) into (37), we get the following approximations:

$$\begin{aligned} P^+_{i-1} \;\simeq\; 1 \;-\; C_{i-1} \int_{-\infty}^{S_{i-1}+\Delta S} & \exp\left\{\frac{\varkappa}{\sigma^2}\,(s - S_{i-1})\left(\theta + \mu T_{i-1} - \frac{s + S_{i-1}}{2}\right)\right\} \\ & \times \left[\exp\left\{-\frac{(s-S_{i-1})^2}{2\sigma^2\,\Delta t}\right\} - \exp\left\{-\frac{(s - S_{i-1} - 2\Delta S)^2}{2\sigma^2\,\Delta t}\right\}\right] ds \,, \end{aligned}$$

where

$$C_{i-1} \;=\; \frac{1}{\sqrt{2\pi\,\Delta t}\,\sigma}\exp\left\{\frac{\varkappa \Delta t}{2}\left(1 - \frac{\varkappa}{\sigma^2}(\theta + \mu T_{i-1} - S_{i-1})^2\right)\right\} \,.$$

Similarly,

$$P_{i-1}^{-} \simeq 1 - C_{i-1} \int_{S_{i-1}-\Delta S}^{+\infty} \exp\left\{\frac{\varkappa}{\sigma^2}(s-S_{i-1})\left(\theta+\mu T_{i-1}-\frac{s+S_{i-1}}{2}\right)\right\} \times \left[\exp\left\{-\frac{(s-S_{i-1})^2}{2\sigma^2\,\Delta t}\right\} - \exp\left\{-\frac{(s-S_{i-1}+2\Delta S)^2}{2\sigma^2\,\Delta t}\right\}\right] ds \ .$$

The integrals in the above expressions for P_{i-1}^{+} and P_{i-1}^{-} can be computed in closed form. We finally get

$$\begin{aligned} P_{i-1}^{+} &\simeq 1 - \frac{\exp(k_0)}{\sqrt{1+\varkappa\,\Delta t}}\left(\exp(k_1)\,\Phi(z_1) - \exp(k_2)\,\Phi(z_2)\right) \ , \\ &\text{where} \qquad (45) \\ k_0 &= \frac{\varkappa\,\Delta t}{2}\left(1-\frac{\varkappa}{\sigma^2}(\theta+\mu T_{i-1}-S_{i-1})^2\right) \ , \\ k_1 &= \frac{\varkappa^2\,\Delta t\,(\theta+\mu T_{i-1}-S_{i-1})^2}{2\,\sigma^2\,(1+\varkappa\,\Delta t)} \ , \\ z_1 &= \frac{\varkappa\,\Delta t\,(S_{i-1}+\Delta S-\theta-\mu T_{i-1})+\Delta S}{\sigma\,\sqrt{\Delta t\,(1+\varkappa\,\Delta t)}} \ , \\ k_2 &= \frac{\varkappa}{2\sigma^2\,(1+\varkappa\,\Delta t)} \\ &\times \left(\varkappa\,\Delta t\,(\theta+\mu T_{i-1}-S_{i-1})^2 - 4(S_{i-1}+\Delta S-\theta-\mu T_{i-1})\,\Delta S\right) \ , \\ z_2 &= \frac{\varkappa\,\Delta t\,(S_{i-1}+\Delta S-\theta-\mu T_{i-1})-\Delta S}{\sigma\,\sqrt{\Delta t\,(1+\varkappa\,\Delta t)}} \ , \end{aligned}$$

and $\Phi(z)$ is the cumulative distribution function (CDF) of the standard normal distribution:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-\frac{x^2}{2}\right)dx \ .$$

Similarly, with the same k_0 and k_1 as above,

$$\begin{aligned} P_{i-1}^{-} &\simeq 1 - \frac{\exp(k_0)}{\sqrt{1+\varkappa\,\Delta t}}\left(\exp(k_3)\,\Phi(-z_3) - \exp(k_4)\,\Phi(-z_4)\right) \ , \\ &\text{where} \qquad (46) \\ k_3 &= k_1 \ , \\ z_3 &= \frac{\varkappa\,\Delta t\,(S_{i-1}-\Delta S-\theta-\mu T_{i-1})-\Delta S}{\sigma\,\sqrt{\Delta t\,(1+\varkappa\,\Delta t)}} \ , \\ k_4 &= \frac{\varkappa}{2\sigma^2\,(1+\varkappa\,\Delta t)} \\ &\times \left(\varkappa\,\Delta t\,(\theta+\mu T_{i-1}-S_{i-1})^2 + 4(S_{i-1}-\Delta S-\theta-\mu T_{i-1})\,\Delta S\right) \ , \\ z_4 &= \frac{\varkappa\,\Delta t\,(S_{i-1}-\Delta S-\theta-\mu T_{i-1})+\Delta S}{\sigma\,\sqrt{\Delta t\,(1+\varkappa\,\Delta t)}} \ . \end{aligned}$$

Note that if $\mu = 0$ and $S_{i-1} = \theta$, then the NSOU process will be completely symmetrical about the initial value, and for any given ΔS, formulas (45) and (46) yield $P_{i-1}^{+} = P_{i-1}^{-}$ as expected.

However, one should bear in mind that the above formulas are approximations only, based on the 3-variate Heat Kernel expansion (37) which is an approximations by itself (unless the stochastic process under consideration is a Brownian Motion with a constant drift). In particular, these formulas are applicable to small values of Δt and ΔS only; they cannot be used

to derive asymptotic behavior of the barrier crossing probabilities for long time intervals or "far-away" barriers. For example, from formulas(45) and (46) we get

$$\lim_{\Delta S \to +\infty} P_{i-1}^{\pm} \;=\; 1 - \frac{\exp(k_0 + k_1)}{\sqrt{1 + \varkappa\,\Delta t}} \;=\; O(\varkappa\,\Delta t)$$

instead of exact zero.

6.3 Explicit Calibration Formulas

In the simplest case of a 1-variate locally-normal transition PDF, the optimal vector $\boldsymbol{q}^* = [\varkappa^*, \theta^*, \mu^*, \sigma^{2*}]$ for the NSOU model can easily be obtained analytically by maximizing the log-likelihood function (6). The result is:

$$\begin{aligned}
\varkappa^* &= \frac{D}{2agk + bh - a^2 h - bg^2 - k^2}\ , \\
\theta^* &= \frac{1}{D}\left(dgk + afk + bch - adh - bfg - ck^2\right)\ , \\
\mu^* &= \frac{1}{D}\left(ack + adg + bf - a^2 f - bcg - dk\right)\ , \\
\sigma^{2*} &= \frac{T}{N}\left[e + \frac{1}{E}\left(2acfk + 2adfg + 2cdgk + bc^2 h + bf^2 + d^2 h\right.\right. \\
&\qquad \left.\left. - a^2 f^2 - c^2 k^2 - d^2 g^2 - 2acdh - 2bcfg - 2dfk\right)\right]\ ,
\end{aligned}$$

where

$$\begin{aligned}
a &= \frac{1}{T}\sum_{i=1}^{N} \tau_i\, S_{i-1}\ , \quad b = \frac{1}{T}\sum_{i=1}^{N} \tau_i\, S_{i-1}^2\ , \quad c = \frac{S_N - S_0}{T}\ , \quad d = \frac{1}{T}\sum_{i=1}^{N} S_{i-1}\,(S_i - S_{i-1})\ , \\
e &= \frac{1}{T}\sum_{i=1}^{N} \frac{(S_i - S_{i-1})^2}{\tau_i}\ , \quad T = \sum_{i=1}^{N} \tau_i\ , \quad f = \frac{1}{T}\sum_{i=1}^{N} T_{i-1}\,(S_i - S_{i-1})\ , \\
g &= \frac{1}{T}\sum_{i=1}^{N} \tau_i\, T_{i-1}\ , \quad h = \frac{1}{T}\sum_{i=1}^{N} \tau_i\, T_{i-1}^2\ , \quad k = \frac{1}{T}\sum_{i=1}^{N} \tau_i\, T_{i-1}\, S_{i-1}\ , \\
D &= dg^2 + fk + ach - dh - afg - cgk\ , \\
E &= k^2 + a^2 h + bg^2 - bh - 2agk\ .
\end{aligned}$$

However, the above estimates can in general be **biased**. In case of a classical OU process ($\mu = 0$), it has been shown in [15] that:

- estimations $\varkappa^*$ and σ^{2*} are indeed biased, and closed-form approximations for their biases have been constructed;
- estimation θ^* is unbiased, but has the variance of the order $O(\varkappa^{-2})$, that is, the variance of θ^* becomes unbounded as $\varkappa \to 0$.

In the general case ($\mu \neq 0$), no published results regarding the NSOU parameter estimation biases exist; in particular, nothing is known about the bias of μ^* itself; this is a subject of future research. In this paper, we will assume that:

- biases of $\varkappa^*$ and σ^{2*} can still be approximated by the formulas presented in [15];

- θ^* remains unbiased, and furthermore, μ^* is unbiased as well; intuitively, the latter assumption can be justified by a similarity between MLE calibration of the θ and μ parameters and least-squares fitting of the line $\theta + \mu(t - t_*)$, which (in the classical case) is unbiased.

Then it follows from the results of [15] that the following estimates for $\varkappa$ and σ^2 which are unbiased up to higher-order terms in N^{-1} and T^{-1}:

$$\begin{aligned} \bar{\varkappa} &= \frac{N}{N+3}\left(\varkappa^* - \frac{4}{T}\right) , \\ \bar{\sigma^2} &= \frac{\sigma^{2*}}{1 - \left(\frac{N}{2} - \frac{17}{12}\bar{\varkappa}T\right)} . \end{aligned}$$

The resulting vector of estimated parameters is:

$$\bar{\boldsymbol{q}} = [\bar{\varkappa},\, \bar{\theta} = \theta^*,\, \bar{\mu} = \mu^*,\, \bar{\sigma^2}] .$$

Finally, we note that generic (model-independent) methods of bias elimination also exist (e.g. bootstrap bias correction [15]) but we do not discuss them here.

6.4 Applications: Reversion Trading Strategy

Therefore, the above NSOU model calibrator acts as a non-liner filter which maps time series of asset prices $\{S_i\}$ into the space of vectors $\bar{\boldsymbol{q}}$. Informally speaking, it is reasonable to expect that the calibrated vectors $\bar{\boldsymbol{q}}$ vary "more slowly" or "more smoothly" over the time than the raw prices S_i. Recall that by "time" we mean either astronomical time or liquidity time (i.e., monotonically-increasing trading volume). Such a filter can be used, for example, in momentum-reversion or market-making trading strategies, for the purpose of statistical continuous price prediction.

For example, let us consider a statistical real-time signal for taking long or short position in the asset under consideration, based on the price reversion property of an asset following the NSOU dynamics. The algorithm is the following:

1. Let S_t be the current price at the current time instant t.
2. Calibrate the NSOU model on the market data at N previous time instants: $S_{t_0}, \ldots, S_{t_N} = S_t$ using one of the methods discussed in this paper, e.g. the fast analytical method. The values of N and the time or liquidity steps τ (say constant for all intervals) could be chosen with the following considerations in mind:
   - τ should be large enough to compensate for effects of market micro-structure, such as the finite granularity of price steps and/or the bid-ask spread, if S_t is actually a mid-price. As a rule of thumb, one could require $\sigma\sqrt{\tau} > \delta$, where δ is the price step or average bid-ask spread;
   - the values of N and $T = N\tau = t_N - t_0$ should ideally be chosen large enough to constrain the standard deviations of the estimated $\bar{\boldsymbol{q}}$ values by some acceptable upper bounds; however, it was pointed out above that the standard deviation of $\bar{\theta}$ can be very large anyway for small $\bar{\varkappa}$; so in practice, one should experiment with multiple values of N, perhaps in the range from 100 to 1000 or above (if long-term reversion effects are of interest) for our 3-parametric NSOU model.
3. Perform bias correction. Let $\bar{\boldsymbol{q}}$ be the resulting vector of model parameters.

4. Check the "physical" validity of the calibrate parameters $\bar{\boldsymbol{q}}$:

   - if we got $\bar{\varkappa} \leq 0$, then there is no reversion in the current market regime, and our model is not applicable at the current time t (though it may become applicable later);
   - also, if we aim at capturing true market reversion effects, and also taking into account that the standard deviation of $\bar{\theta}$ may be very large, it may be a good idea to perform the following "sanity check" on $\bar{\theta}$:

$$\min_{t_0<t'<t_N} S_{t'} < \bar{\theta} < \max_{t_0<t'<t_N} S_{t'} \ ;$$

   - needless to say, we must always have $\bar{\sigma^2} > 0$.

   If any of the above checks fail, do not enter any position based on this model at time t.

5. Perform the statistical model validation procedure (a Variance Ratio test on the residual Random Walk as discussed in Section 5). Again, if this test fails (i.e. we were able to falsify the hypothesis that the residual process is a Random Walk at a given confidence level), the model is not applicable, so refrain from taking any positions.

6. Otherwise, the NSOU model is considered to be applicable at the current time t. Specify the target gross profit (before any transaction costs) $\Delta S > 0$, the maximum time horizon $\Delta t > 0$ for realizing that profit, and the target probability α of profit realization.

7. Then compute the probabilities P_N^+ and P_N^- using formulas (45) and (46) with $i - 1 = N$ and with the estimated parameter vector $\bar{\boldsymbol{q}}$. These quantities are the probabilities of realizing profit ΔS on taking Long or Short position, resp., at the current time $t = t_N$.

8. If $P_N^+ \geq \alpha$, enter Long position; if $P_N^- \geq \alpha$, enter Short position. Note that it is possible that both Long and Short positions are entered (e.g. passively, in the course of market-making), or none of them.

9. Exit the position(s) either by reaching the take-profit target ΔS or by incurring a certain stopping loss ΔL.

10. "Hyper-parameters" τ, N, Δt, ΔS, α, ΔL can be calibrated using historical back-testing of the trading strategy, optimizing the P&L, Sharpe Ratio or other performance criteria.

7 Conclusions

We have proposed an innovative method for calibrating 1-dimensional stochastic models of type (1), which occur frequently in various problems of Quantitative Finance. The algorithm of the method is given by (38). The distinctive features of the method are the following:

- the method allows us to calibrate the model parameters belonging to both the trend (drift) term and the volatility term, via a single unified algorithm, for a variety of asset classes;
- the method could therefore be applied to multiple problems in Quantitative Finance, in particular to Algorithmic Trading and Options Pricing;
- the method makes use of intervalized OHLC market data of the underlying asset (with intervals constructed either over a wall clock, or over a liquidity clock), but it does not require any Option prices in order to operate;

- the method is based on the modern Heat Kernel theory, which is used to construct multivariate transition densities within the MLE framework;
- however, MLE results can in general be biased, and/or may have large variances; various methods (both model-specific and generic) for bias elimination are known from the literature, and should be applied in conjunction with the calibration methods presented here;
- the validity of a given model with a calibrated set of parameters may be tested (and falsified) by applying a Variance Ratio test to the residual random walk process;
- using the method of 3-variate Heat Kernels, formulas for barrier crossing probabilities of the NSOU process have been derived;
- an example of a reversion trading strategy based on the developed quantitative framework was given.

References

[1] Yacine Aït-Sahalia. "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach". In: *Econometrica* 70.1 (2002), pp. 223–262. URL: https://www.princeton.edu/~yacine/mle.pdf.

[2] Ivan G. Avramidi. *Heat Kernel Method and its Applications*. Cham: Birkhäuser, 2015. ISBN: 978-3319262659.

[3] Damiano Brigo and Fabio Mercurio. *Interest Rate Models — Theory and Practice: With Smile, Inflation and Credit*. 2nd. Berlin: Springer, 2006. ISBN: 978-3540221494.

[4] Amélie Charles and Olivier Darné. *Variance Ratio Tests for Random Walk: An Overview*. 2013. URL: https://hal.archives-ouvertes.fr/hal-00771078.

[5] Willa W. Chen and Rohit S. Deo. "The Variance Ratio Statistic at Large Horizons". In: *Econometric Theory* 22.2 (2006), pp. 206–234. URL: https://www.jstor.org/stable/4093223.

[6] Hélyette Geman. *Commodities and Commodity Derivatives: Modeling and Pricing for Agriculturals, Metals and Energy*. Chichester: Wiley, 2005. ISBN: 978-0470012185.

[7] Hua He, William P. Keirstead, and Joachim Rebholz. *Double Lookbacks*. Finance Working Paper #248. Walter A. Haas School of Business, University of California Berkeley, May 1995. URL: http://www.haas.berkeley.edu/groups/finance/WP/rpf248.pdf.

[8] Erich Kamke. *Differentialgleichungen Lösungsmethoden und Lösungen*. 10th. Vol. I. Reprinted 2012. Wiesbaden: Springer, 1977. ISBN: 978-3663059264.

[9] Mathieu Kessler. "Estimation of an Ergodic Diffusion from Discrete Observations". In: *The Scandinavian Journal of Statistics* 24 (1997), pp. 211–229. URL: https://perso.univ-rennes1.fr/jian-feng.yao/gdt/docs/Kessler97.pdf.

[10] Vladimir A. Marchenko. *Sturm-Liouville Operators and Applications*. Revised. Providence, RI: AMS Chlsea Publishing, 2011. ISBN: 978-0821853160.

[11] Peter Molnár. "Properties of Range-Based Volatility Estimators". In: *International Review of Financial Analysis* 23 (2012), pp. 20–29. URL: https://www.sciencedirect.com/science/article/pii/S1057521911000731.

[12] Louis Paulot. *Asymptotic Implied Volatility at the Second Order With Application to the SABR Model*. 2009. URL: https://arxiv.org/pdf/0906.0658.pdf.

[13] Marcos López de Prado. *The 10 Reasons Most Machine Learning Funds Fail*. 2017. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3104816.

[14] Steven E. Shreve. *Stochastic Calculus for Finance*. Vol. II: Continuous-Time Models. New York: Springer, 2004. ISBN: 978-0387401010.

[15] Chengyong Tang. "Parameter Estimation and Bias Correction for Diffusion Processes and a Nonparametric Approach to Census Population Size Estimation". PhD thesis. Ames, Iowa: Iowa State University, 2008. URL: https://lib.dr.iastate.edu/rtd/15803.

[16] Paul Wilmott, Jeff Dewynne, and Sam Howison. *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press, 1994. ISBN: 978-0952208204.