1 Introduction

We consider two-dimensional correlated Brownian motion with absorbing boundaries:

$$X(t) = x_0 + \mu_x t + \sigma_x W_x(t) \qquad a_x < X(t) < b_x \tag{1}$$
$$Y(t) = y_0 + \mu_y t + \sigma_y W_y(t) \qquad a_y < Y(t) < b_y \tag{2}$$

where W_i are standard Brownian motions with $\text{Cov}(W_1(t), W_2(t)) = \rho t$ for $0 < t' \leq t$. In particular, we find the joint transition density function for $(X(t), Y(t))$ under the boundary conditions:

$$\Pr\left(X(t) \in dx, Y(t) \in dy, \min_{t'} X(t') \geq a_x, \max_{t'} X(t') \leq b_x, \min_{t'} Y(t') \geq a_y, \max_{t'} Y(t') \leq b_y | X(0) = x_0, Y(0) = y_0, \theta\right), \tag{3}$$

with $\theta := (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. This function, which we shorten to $q(x,y,t)$ from now on, is the solution to the Fokker-Planck equation [Oksendal, 2013]:

$$\frac{\partial}{\partial t} q(x,y,t') = -\mu_x \frac{\partial}{\partial x} q(x,y,t') - \mu_y \frac{\partial}{\partial y} q(x,y,t') + \frac{1}{2}\sigma_x^2 \frac{\partial^2}{\partial x^2} q(x,y,t') + \rho\sigma_x\sigma_y \frac{\partial^2}{\partial x \partial y} q(x,y,t') + \frac{1}{2}\sigma_y^2 \frac{\partial^2}{\partial y^2} q(x,y,t'), \tag{4}$$

$$q(a_x, y, t') = q(b_x, y, t') = q(x, a_y, t') = q(x, b_y, t') = 0, \qquad 0 < t' \leq t. \tag{5}$$

Differentiating $q(x,y,t)$ with respect to the boundaries produces the transition density of a particle beginning and ending at the points $(X_1(0), X_2(0))$ and $(X_1(t), X_2(t))$, respectively, while attaining the minima a_x/a_y and maxima b_x/b_y in each coordinate direction:

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q(x,y,t) =$$

$$\Pr\left(X(t) \in dx, Y(t) \in dy, \min_{t'} X(t') = a_x, \max_{t'} X(t') = b_x, \min_{t'} Y(t') = a_y, \max_{t'} Y(t') = b_y \,\middle|\, 0 < t' \leq t, X(0) = x_0, Y(0) = y_0, \theta\right). \tag{6}$$

The transition density (3) with less than all four boundaries has been used in computing first passage times [Kou et al., 2016, Sacerdote et al., 2016], with application to structural models in credit risk and default correlations [Haworth et al., 2008, Ching et al., 2014]. He et al. [1998] use variants of (6) with respect to some of the boundaries to price financial derivative instruments whose payoff depends on **some** of the observed maxima/minima. Horst et al. [2012] use a full likelihood-based (Bayesian) approach to estimate volatility in *univariate* financial timeseries where open, closing, highest, and lowest prices are included. Their work fits into a body of literature and collection of techniques by practitioners where the observed range of prices is used to make similar estimates. In this paper we will provide the efficient numerical methods necessary to carry out similar inferential procedures with correlated financial timeseries.

Closed-form solutions to (4) - (5) are available for some parameter regimes. When $\rho = 0$, the transition density of the process is the solution to a well-understood Sturm-Liouville problem where the eigenfunctions of the differential operator are sine functions. When $a_1 = -\infty$ and $b_1 = \infty$, the method of images can be used to enforce the remaining boundaries. For either $a_1, a_2 = -\infty$ or $b_1, b_2 = \infty$, eigenfunction of the Fokker-Plank equation can be found in radial coordinates. Both of these techniques are used and detailed by He et al. [1998]. However, to the best of our knowledge, there is no closed-form solution to the general problem in (4) - (5). This also limits the available ways to compute (6), with the most straightforward approach being finite difference with respect to the boundary conditions. This, however, requires one to solve at least 16 eigenvalue problems to evaluate the density function for a single observation (see Section 2.1.1), motivating the need for an efficient numerical method to solve (4) - (5).

It is still possible to approach the general problem by proposing a biorthogonal expansion in time and space (Risken [1989], sections 6.2), where the eigenfunctions for the differential operator are approximated as a sine series satisfying the boundary conditions. However, a drawback of this out-of-the-box solution is that the system matrix for the corresponding eigenvalue problem is dense. Additionally, for nonzero ρ, it is necessary to have a large number of basis elements for an accurate approximation. The denseness and size of the resultant system matrices makes the eigenvalue expansion a slow, if not unfeasible, solution. An alternative is to use a finite difference scheme to directly solve the evolution problem after suitable transformations. However, both of these methods need a high degree of numerical resolution to produce practically useful approximations of the transition density. We conjecture that these inefficiencies come from either using a *separable* representation for the differential operator (trigonometric series) or introducing numerical diffusion (finite difference).

In this paper, we propose a solution to the general problem (4) - (5) which is obtained by combining a small-time analytic solution with a finite-element method. Our method directly takes into account the correlation parameter present in the differential operator in order to efficiently represent the analytic small-time solution and propagate it forward in time. We apply our computational method to estimate equation parameters with a maximum likelihood approach in settings where the model assumptions of constant $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ and Brownian motion driving stochastic evolution are appropriate. Section 2 outlines some methods we considered, including the one we used. Section [] includes our numerical experiments.

2 Approximate Numerical Solutions

Before considering solutions to the full Fokker-Planck equation (4) - (5), we simplify the PDE by proposing a scaling transformation and an exponential decomposition of the solution, so that we can construct

$$p(x,y,t) = \exp(\alpha x + \beta y + \gamma t) q\left(x(b_x - a_x) + a_x, \quad y(b_y - a_y) + a_y, \quad t\right),$$

where

$$\begin{aligned}
\alpha &= -\frac{\mu_x}{\sigma_x^2} - \frac{\rho}{\sigma_x \sigma_y (1-\rho^2)} \left(-\frac{\mu_y}{\sigma_y^2} + \frac{\mu_x \rho}{\sigma_x \sigma_y} \right), \\
\beta &= \left(-\frac{\mu_y}{\sigma_y^2} + \frac{\mu_x \rho}{\sigma_x \sigma_y} \right), \\
\gamma &= \frac{1}{2} \left(\frac{\sigma_x}{(b_x - a_x)} \right)^2 \alpha^2 + \frac{1}{2} \left(\frac{\sigma_y}{(b_y - a_y)} \right)^2 \beta^2 + \alpha\beta.
\end{aligned}$$

This new formula satisfies the simpler diffusion equation:

$$\frac{\partial}{\partial t} p(x,y,t) = \mathcal{L} p(x,y,t), \quad (x,y) \in (0,1) \times (0,1) := \Omega \tag{7}$$

subject to the constraints

$$\begin{aligned}
&p(x,y,t) = 0, && \text{for } (x,y) \in \{x | x = 0\} \cup \{y | y = 0\}, \\
&p(x,y,0) = \delta\left(x - \frac{x_0 - a_x}{b_x - a_x} \right) \delta\left(y - \frac{y_0 - a_y}{b_y - a_y} \right),
\end{aligned}$$

where the differential operator $\mathcal{L}$ takes the form

$$\mathcal{L} = \frac{1}{2} \tau_x^2 \frac{\partial^2}{\partial x^2} + \rho \tau_x \tau_y \frac{\partial^2}{\partial x \partial y} + \frac{1}{2} \tau_y^2 \frac{\partial^2}{\partial y^2},$$

where $\tau_x = \frac{\sigma_x}{(b_x - a_x)}$, $\tau_y = \frac{\sigma_y}{(b_y - a_y)}$. Note here that under this transformation ρ remains the same as in the original problem. We will call equation (7) the *normalized* problem and will consider its solution in terms of the diffusion parameters (τ_x, τ_y, ρ) without a loss of generality.

2.1 Eigenfunction Expansion

Following Section 6.2 of Risken [1989], we may use the biorthogonal decomposition of the solution as a sum of eigenfunctions and time-dependent coefficients determined by eigenvalues:

$$p(x,y,t) = \phi_\nu(x,y)e^{-\lambda_\nu t}, \tag{8}$$

where the eigenfunctions $\phi_\nu(x,y)$ satisfy the boundary conditions. Because the differential operator $\mathcal{L}$ is self-adjoint, the family of eigenfunctions is complete in the Hilbert space $L_2(\Omega)$. Moreover, the eigenvalues are bounded below by a positive constant c, so that the solution behaves as expected (see section 6.3 of Risken [1989]).

Since we require $\phi_\nu(x,y)$ to be zero on the boundaries, we may represent the eigenfunction as a linear combination of sines

$$\phi_\nu(x,y) = \sum_{l=0}^{L}\sum_{m=0}^{M} c_{l,m,\nu} \sin(2\pi l x)\sin(2\pi m y) := \Psi(x,y)^T c_\nu,$$

where we have truncated the infinite series for some suitably large L and M and defined

$$\begin{aligned}
\psi_{l,m}(x,y) &= \sin(2\pi l x)\sin(2\pi m y), \\
\Psi(x,y) &= (\psi_{0,0}(x,y), \ldots, \psi_{L,M}(x,y))^T, \\
c_\nu &= (c_{0,0,\nu}, \ldots, c_{L,M,\nu})^T.
\end{aligned}$$

The biorthogonal representation (8) leads to the eigenvalue problem

$$\mathcal{L}\phi_\nu = -\lambda_\nu \phi_\nu, \tag{9}$$

where $\mathcal{L}$ is the differential operator in the normalized Fokker-Planck equation. Applying $\mathcal{L}$ to ϕ_ν produces the linear system

$$\mathcal{L}\phi_\nu = \mathcal{L}(\Psi(x,y)^T c_\nu) = \mathcal{L}(\Psi(x,y)^T)c_\nu = (A\Psi(x,y))^T c_\nu,$$

where A is a constant matrix dependent on $\tilde{\theta}$. In the case where $\rho = 0$, A is diagonal because $\{\psi_{l,m}(x,y)\}_{l,m}$ are the eigenfunctions to $\mathcal{L}$. When $\rho \neq 0$, A is no longer diagonal and is in fact dense. This caused by the mixing terms

$$\frac{\partial^2}{\partial x \partial y}\sin(2\pi l x)\sin(2\pi m y) = (2\pi l)(2\pi m)\cos(2\pi l x)\cos(2\pi m y)$$

being the products of cosine functions, which have an inefficient sine series representation [CITE]. Substituting the linear representation of $\mathcal{L}\phi_\nu$ into the eigenvalue problem (9), we arrive to the system

$$\Psi(x,y)^T A^T c_\nu = -\lambda_\nu \Psi(x,y)^T c_\nu \quad \Leftrightarrow \quad A^T c_\nu = -\lambda_\nu c_\nu$$

whose solution gives the family of orthonormal eigenfunctions. As mentioned already, the efficiency of this approach is dependent on the cost of solving the eigenvalue problem $A^T c_\nu = -\lambda_\nu c_\nu$.

With all of the eigenpairs (c_ν, λ_ν), the approximate solution is then

$$p(x,y,t) \approx \sum_\nu \Psi(x,y)^T c_\nu e^{-\lambda_\nu t}.$$

2.1.1 Likelihood Calculation

Recalling equation (6), to compute the likelihood we are interested in we must take derivatives of the solution $p(x,y,t)$ with respect to the boundary parameters (a_x, b_x, a_y, b_y). Because the eigenvalues and eigenvectors for this problem are functions of the parameters without explicit analytic form, $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x,y,t)$ must be computed numerically using a finite difference approximation. We must therefore find the eigenvalues and eigenvectors for each of the sixteen perturbed problems

$$p(x,y,t|a_x \pm \varepsilon, b_x \pm \varepsilon, a_y \pm \varepsilon, b_y \pm \varepsilon),$$

such that

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x,y,t) \approx \frac{\sum_k c_k p(x,y,t|a_x \pm \varepsilon, b_x \pm \varepsilon, a_y \pm \varepsilon, b_y \pm \varepsilon)}{(2\varepsilon)^4}.$$

Because a good, working approximation to $p(x,y,t)$ requires many terms in the expansion of the eigenfunctions, and because each resultant system matrix is dense, repeated computation of the likelihood in this way is unfeasible.

2.2 Finite Difference

A finite difference method which approximates the spatial derivatives in problem (7) requires the solution a system of differential equations

$$\dot{c}(t) = Ac(t) \Rightarrow \qquad c(t) = \exp(At)\, c(0), \tag{10}$$

which reduces to the eigenvalue decomposition of a matrix A. Here, $c(t)$ is a vector of coefficients representing the value of the solution in (7) on each point on a grid over Ω at time t, and the product $Ac(t)$ approximates $\mathcal{L}p(x,y,t)$. The system matrix A is dependent on the discretization scheme used to approximate $\mathcal{L}$. Using a central in space scheme over a **regular** grid on Ω with $\Delta x = \Delta y = h$,

$$A = \frac{1}{2}\tau_x^2 \frac{1}{h^2} A_{x,x} + \rho\tau_x\tau_y \frac{1}{4h^2} A_{x,y} + \frac{1}{2}\tau_y^2 \frac{1}{h^2} A_{y,y},$$

where each of the matrices $A_{x,x}, A_{x,y}, A_{y,y}$ approximate $\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x \partial y}, \frac{\partial^2}{\partial y^2}$, respectively. For example, using the labeling $c_{l,m}(t) \to c_k(t)$ on the vector $c(t)$ to denote the approximation of the solution at point (x_l, y_m) on the grid,

$$\frac{1}{2}\tau_x^2 \frac{\partial^2}{\partial x^2} p(x_l, y_m, t) \approx \frac{1}{2}\tau_x^2 \left(\frac{c_{l+1,m}(t) - 2c_{l,m}(t) + c_{l-1,m}(t)}{h^2} \right) = \frac{1}{2}\tau_x^2 \frac{1}{h^2} A_{x,x}[k,]c(t)$$

where $A_{x,x}[k,]$ is the k^{th} row of $A_{x,x}$. Similarly,

$$\rho\tau_x\tau_y \frac{\partial^2}{\partial x \partial y} p(x_l, y_m, t) \approx \rho\tau_x\tau_y \left(\frac{c_{l+1,m+1}(t) - c_{l+1,m-1}(t) - c_{l-1,m+1}(t) + c_{l-1,m-1}(t)}{4h^2} \right) = \rho\tau_x\tau_y \frac{1}{4h^2} A_{x,y}[k,]c(t)$$

It should be noted here that a regular grid approach with a constant h is appealing, because it allows us to construct once and store the matrices $A_{x,x}, A_{x,y}, A_{y,y}$, which saves valuable computational resources if we are to solve the finite difference eigenproblem (10) repeatedly for different parameter values (τ_x, τ_y, ρ).

Unlike the system matrix for the trigonometric expansion, the system matrix A here is sparse: each row of $(A_{x,x}, A_{x,y}, A_{y,y})$ is composed of all zeros except for three or four entries. This structure does not change as $h \to 0$. The eigenvalue problem is therefore much cheaper to solve. The system matrix can be made even sparser on a regular grid by performing a 45° rotation which removes the mixing term $\rho\tau_x\tau_y \frac{\partial^2}{\partial x \partial y}$ from the problem PDE and preserves the boundaries.

However, the fundamental limitation of using a finite difference method is that differentiation with respect to boundaries is also done using finite differences and it is of fourth order. Given this high order differentiation, h cannot be made too small because truncation error becomes an issue relatively quickly. More explicitly, assuming consistency and stability of the finite difference approximation to $\mathcal{L}$, we can write down the finite difference solution as

$$p_{FD}(x_l, y_m, t|a_x) = p(x_l, y_m, t) + O(h^2; a_x).$$

Note here the residual term $O(h^2; a_x)$ a function of the paramter a_x. The finite difference approximation yields:

$$\begin{aligned}\frac{p_{FD}(x_l, y_m, t|a_x) - p_{FD}(x_l, y_m, t|a_x - \varepsilon)}{\varepsilon} &= \frac{p(x_l, y_m, t|a_x) - p(x_l, y_m, t|a_x - \varepsilon)}{\varepsilon} + \frac{O(h^2|a_x) - O(h^2|a_x - \varepsilon)}{\varepsilon} \\ &\approx \frac{\partial}{\partial a_x} p(x_l, y_m, t) + O(\varepsilon) + \frac{O(h^2|a_x) - O(h^2|a_x - \varepsilon)}{\varepsilon}.\end{aligned}$$

If the residual term $O(h^2|a_x)$ is differentiable with respect to a_x,

$$\frac{O(h^2|a_x) - O(h^2|a_x - \varepsilon)}{\varepsilon} \to O(h^2) \text{ as } \varepsilon \to 0.$$

Otherwise

$$\frac{O(h^2|a_x) - O(h^2|a_x - \varepsilon)}{\varepsilon} \to O(h^2/\varepsilon^r) \text{ as } \varepsilon \to 0.$$

for some $r > 0$. Under this condition, as $\varepsilon \to 0$ with h fixed, the term $O(h^2/\varepsilon^r)$ begins to dominate the approximation. Hence, for a fixed h, we have a bound on how accurately we may represent the likelihood function provided the finite difference residual term is not differentiable with respect to the boundary parameters.

The finite value for h introduces yet another practical concern. On a regular grid, the delta function initial condition needs to be either rounded to the nearest grid point or represented as a weighted sum of delta functions on the four nearest grid points. Either approach introduces a numerical diffusion into the problem which, for a finite h, can bias the numerical solution. We will demonstrate this phenomenon in Section [].

2.3 Semidiscrete Galerkin Method

We propose a semidiscrete Galerkin (continuous in time, discrete in space) solution to the general diffusion problem (7). The success of this approach is based on an analytic approximation for the solution $p(x, y, t)$ for small time t and a basic convergence estimate from approximation theory for semidiscrete Galerkin-type solutions to parabolic problems.

Our numerical method produces a functional representation for the approximate solution which 1) imposes a computational burden comparable to or better than that of the finite difference method and 2) is infinitely differentiable with respect to the boundary parameters, allowing us to perform the crucial numerical differentiation with respect to these parameters.

As described in Section 2.1, the solution to the model problem has the eigenfunction expansion

$$p(x, y, t) = \sum_{\nu=0}^{\infty} \phi_\nu(x, y) e^{-\lambda_\nu t}.$$

Proceeding with the standandard Galerkin approach, we propose a solution $p_k(x, y, t)$ of similar form

$$p_k(x, y, t) = \sum_{i=0}^{k} c_i(t) \psi_i(x, y),$$

where the basis functions $\psi_i(x,y)$ satisfy the boundary conditions. We also require that all first- and second-order derivatives of $\psi_i(x,y)$ are in $L_2(\Omega)$, i.e. $\psi_k(x,y) \in W_2^2(\Omega)$. In other words, we look for a solution in the span of $S_k = \{\psi_0, \ldots, \psi_k\}$. Since p_k is an approximation to the solution p, it does not follow the differential equation exactly nor can it represent the initial condition fully. We capture this by defining residuals

$$\begin{aligned} \frac{\partial}{\partial t} p_k(x,y,t) - \mathcal{L} p_k(x,y,t) &:= R_e(k), \\ p(x,y,0) - p_k(x,y,0) &:= R_0(k). \end{aligned}$$

There are various conditions that could be imposed on the residual functions (see Section 2.10.3 of Norrie and De Vries [1973] for a summary). The *orthogonality* condition coincides with the Galerkin procedure:

$$\int_\Omega R_e(k)\psi_i(x,y)dxdy = 0, \qquad \int_\Omega R_0(k)\psi_i(x,y)dxdy = 0, \qquad i = 0, \ldots k, \tag{11}$$

which is equivalent to the weak formulation of the heat problem

$$\begin{aligned} (\partial_t p_k(x,y,t), \psi_i) &= (\mathcal{L} p_k(x,y,t), \psi_i), \\ (p_k(x,y,0), \psi_i) &= (p(x,y,0), \psi_i), \end{aligned}$$

where $(\cdot,\cdot)$ is the usual inner product in $L_2(\Omega)$. The orthogonality conditions (11) lead to the system of equations

$$\begin{aligned} M\dot{\mathbf{c}}(t) &= S\mathbf{c}(t), \\ M\mathbf{c}(0) &= \mathbf{p}(x,y), \end{aligned}$$

where M is the mass matrix, S is the stiffness matrix, and $\mathbf{p}(x,y,0)$ is the vector projection of $p(x,y,0)$ onto the span of S_k with elements

$$\begin{aligned} M_{ij} &= \int_\Omega \psi_i \psi_j dxdy, \\ S_{ij} &= -\frac{1}{2}\tau_x^2 \int_\Omega \left(\frac{\partial}{\partial x}\psi_i(x,y)\right)\left(\frac{\partial}{\partial x}\psi_j(x,y)\right)dxdy - \rho\tau_x\tau_y \int_\Omega \left(\frac{\partial}{\partial x}\psi_i(x,y)\right)\left(\frac{\partial}{\partial y}\psi_j(x,y)\right)dxdy \\ &\quad -\frac{1}{2}\tau_y^2 \int_\Omega \left(\frac{\partial}{\partial y}\psi_i(x,y)\right)\left(\frac{\partial}{\partial y}\psi_j(x,y)\right)dxdy \\ \mathbf{p}_i(x,y) &= \int_\Omega p(x,y,0)\psi_i(x,y)dxdy. \end{aligned}$$

The entries of S_{ij} are computed with integration by parts and using the boundary conditions. The semidiscrete Galerkin approximation becomes

$$p_k(x,y,t) = \psi(x,y)^T \exp\left(M^{-1}St\right)\mathbf{c}(0),$$

with $\psi(x,y) = (\psi_0(x,y), \ldots, \psi_k(x,y))^T$.

A bound on the closeness of the approximate solution $p_k(x,y,t)$ to the strong solution $p(x,y,t)$ is developed in Bramble et al. [1977] shows that the Galerkin approximation we use converges to the strong solution, and it motivates the thrust of our numerical solution. First, we define the *error* term

$$e_k(t) = p(x,y,t) - p_k(x,y,t),$$

as well as the norm

$$\|w\|_2 = \sum_{j=0}^{\infty} \lambda_j^2 (w, \phi_j)$$

for the eigenpairs (λ_j, ϕ_j) of the operator $\mathcal{L}$. As referred in Bramble et al. [1977], functions $w \in L_2(\Omega)$ with $\|w\|_2 < \infty$ are also in $W_2^2(\Omega)$. In a sense, $\|\cdot\|_2$ measures the variation of functions. Functions having higher-order eigenmodes have a bigger norm, and vice versa. Finally, if we have the condition (corresponding to equation 2.1 in Bramble et al. [1977])

$$\|p(x,y,0) - p_k(x,y,0)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x,y,0)\|_2. \tag{12}$$

as a decreasing function $h(k)$, Theorem 2.1 in Bramble et al. [1977] applies and we have the error estimate

$$\|e(t)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x,y,0)\|_2. \tag{13}$$

We can ensure condition (12) is met if S_k is complete in $L_2(\Omega)$ as k grows. The other two conditions for Theorem 2.1 to apply are demonstrated by Bramble et al. [1977] for the Galerkin method. Equation (13) can be summarized in a simple way: the error of the method is controlled by how much variation the initial condition has; the rate of decrease of the error is controlled by how well the span of S_k represents the initial condition compared to the variation of the initial condition.

However, the initial condition for (7) requires all in eigenmodes being included in the representation of the initial conditions, so that

$$\|p(x,y,0)\|_2 = \left\| \delta\left(x - \frac{x_0 - a_x}{b_x - a_x}\right) \delta\left(y - \frac{y_0 - a_y}{b_y - a_y}\right) \right\|_2 = +\infty.$$

It is obvious that we cannot apply the error estimates above. However, evolving the solution forward in time, even for a short period, diffuses the delta-function IC and attenuates out the highest frequency modes. Further, because solutions to the diffusion equation are smooth, $p(x,y,t_\varepsilon)$ will give us a smooth initial function which will make the error bounds admissible.

2.3.1 Small-time Solution

The small-time solution is derived by considering the fundamental solution $G(x,y,t)$ for the unbounded problem in (7), which is the bivariate Gaussian density with mean and covariance determined by the initial condition and the diffusion parameters [Stakgold and Holst, 2011]. We can then find a small enough t_ε such that $G\left(x,y,t_\varepsilon \middle| \frac{x_0-a_x}{b_x-a_x}, \frac{y_0-a_y}{b_y-a_y}\right)$ is numerically zero on the boundaries of Ω. This is done in the following way.

1. Scale and rotate the coordinate axes by

   $$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \text{Transformation goes here} \begin{pmatrix} x \\ y \end{pmatrix}$$

   so that the problem obeys the diffusion equation. The computational domain is transformed to the parallelogram $\tilde{\Omega}$. The transformed initial condition will be denoted as (ξ_0, η_0).

2. The fundamental solution $G(\xi,\eta,t|\xi_0,\eta_0)$ in this coordinate frame which does not take into account boundaries follows the bivariate Gaussian probability density function

   $$G(\xi,\eta,t|\xi_0,\eta_0) = \frac{1}{2\pi t} \exp\left(-\frac{(\xi-\xi_0)^2 + (\eta-\eta_0)^2}{2t}\right).$$

   We define **distance** between $G(\xi,\eta,t|\xi_0,\eta_0)$ and any of the linear boundaries of $\tilde{\Omega}$ as the Euclidean norm of the segment of the vector extending from (ξ_0,η_0) and normal to the boundary. There are four such distances d_1, d_2, d_3, d_4. Assume that they are listed in increasing magnitude.

   Setting $t_\varepsilon = d_2/8$ ensures that the fundamental solution $G(\xi,\eta,t_\varepsilon|\xi_0,\eta_0)$ is *at most* $\approx 10^{-15}$ on the second-farthest boundary, as well as the other two farthest boundaries. In this way, $G(\xi,\eta,t_\varepsilon|xi_0,\eta_0)$ satisfies the boundary condition on the three farthest boundaries numerically.

3. Reflect the point $(\xi_0, \eta_0) \to (\xi_0', \eta_0')$ about the closest boundary. The image function $G(\xi, \eta, t_\varepsilon | \xi_0', \eta_0')$ satisfies the diffusion equation and is equal to $G(\xi, \eta, t_\varepsilon | \xi_0, \eta_0)$ on the closest boundary. Further, $G(\xi, \eta, t_\varepsilon | \xi_0', \eta_0')$ takes on values less than 10^{-15} on all other boundaries, because it is outside of $\bar{\Omega}$.

   Considering the difference of the two images, the small-time solution

$$p(\xi, \eta, t_\varepsilon) := G(\xi, \eta, t_\varepsilon | \xi_0, \eta_0) - G(\xi, \eta, t_\varepsilon | \xi_0', \eta_0')$$

   satisfies all of the boundaries numerically and also satisfies the governing diffusion equation. Performing a change of variables produces the small-time solution $p(x, y, t_\varepsilon)$.

Using Theorem 5.E of Zeidler [1995], we can solve for $p(x, y, t)$ by considering the smooth $p(x, y, t_\varepsilon)$ as an initial condition and evolving it forward in time by $t - t_\varepsilon$. The bigger t_ε, the less eigenmodes present in $p(x, y, t_\varepsilon)$, and the more accurate our weak solution according to the error estimate

$$\|e(t)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x, y, t_\varepsilon)\|_2.$$

2.3.2 Orthonormal Basis Family

An upper bound of the rate at which the Galerkin approximation converges to $p(x, y, t)$ is given by condition (12), namely by how well the initial condition may be approximated via a projection onto S_k. We motivate the construction of the orthonormal basis functions by once again considering the fundamental solution for the unbounded problem (7). In the absence of boundaries, (7) is solved by the function

$$G(x, y, t | x_0', y_0') = \frac{1}{2\pi\, t\, \tau_x \tau_y \sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x-x_0')^2}{\tau_x^2} - 2\rho \frac{(x-x_0')(y-y_0')}{\tau_x \tau_y} + \frac{(y-y_0')^2}{\tau_y^2} \right) \right\},$$

with $x_0' = (x_0 - a_x)/(b_x - a_x)$, $\quad y_0' = (y_0 - a_y)/(b_y - a_y)$. We choose the family of basis functions $S_k = (\psi_0(x, y), \ldots, \psi_k(x, y))$

$$\psi_i(x, y) = \frac{1}{2\pi\sigma^2 \sqrt{1-\tilde{\rho}^2}} \exp\left\{ -\frac{1}{2(1-\tilde{\rho}^2)\sigma^2} \left((x-x_i)^2 - 2\tilde{\rho}(x-x_i)(y-y_i) + (y-y_i)^2 \right) \right\} x(1-x)\, y(1-y) \qquad (14)$$

for some parameters $(\tilde{\rho}, \sigma)$ and a collection of points $\{(x_i, y_i)\}_{i=0}^k$ which form a grid over Ω. Essentially, the collection $\{\psi_i(x, y | x_i, y_i, \tilde{\rho}, \sigma)\}_{i=0}^k$ is composed of fundamental solutions to a heat diffusion problem tuned by σ and $\tilde{\rho}$, tampered such that their support is on Ω, zero on the boundaries, centered on some grid over Ω, and still smooth. In this manner, our basis function choice is in keeping with the golden rule for the rate of convergence: **The smoother the solution to the original problem and the smoother the functions in the basis space, the faster the convergence of the Ritz-Galerkin method.** (Remark 1(c) in Chapter 2 of Zeidler [1995]). Moreover, all of the basis functions are infinitely differentiable with respect to the boundary and diffusion parameters, so that $\frac{\partial^4 p_k(x,y,t)}{\partial a_x \partial b_x \partial a_y \partial b_y}$ exists and can be computed numerically using finite differences.

The other critical advantage of these elements is that they better resolve the small-time solution $p(x, y, t_\varepsilon)$ by taking into account correlation in the covariance of each kernel. [**if not prove, demonstrate. I'm running out of steam on this.**].

3 Estimation

Consider the problem of estimating the parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ from an i.i.d. sample $(Z_1, \ldots, Z_n)$ from the process in (1) - (2), where each Z_i is the vector of random variables

$$Z_i = (X_i(t), Y_i(t), A_{x,i}, B_{x,i}, A_{y,i}, B_{y,i}).$$

We say that Z_i is sampled from the distribution corresponding to the probability density function (6)

$$Z_i \sim F(\theta),$$

where the distribution F has the usual interpretation

$$F(z=(x,y,a_x,b_x,a_y,b_y)|\theta) = \Pr\left(X(t)\le x, Y(t)\le y, \min_{t'\le t} X(t') \le a_x, \max_{t'\le t} X(t') \le b_x, \min_{t'\le t} Y(t') \le a_y, \max_{t'\le t} Y(t') \le b_y\right).$$

This estimation problem is of particular importance in quantitative finance where the model equations (1) - (2) (with various bells and whistles attached) are widely used. However, to the best of our knowledge, all current *likelihood* methods in the literature either ignore the observed maximum/minimum information or use some of it. Likelihood-free approaches, like that of Rogers and Satchell [1991], on the other hand suffer from not being able to be easily integrated into a bigger inferential framework.

Since we do not have a closed-form solution for the likelihood, we will use an iterative derivative-free maximization algorithm (the Nelder-Mead method; see Lagarias et al. [1998] for review and convergence properties) which requires repeated evauluation of the likelihood. For moderate to large samples sizes this is feasible, because our numerical method is specifically designed for computational efficiency for repeatedly evaluating the density function (6). The maximum likelihood estimator (MLE) for the true parameters based on n samples, which we call $\hat{\theta}_n := (\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$, is especially useful in practical settings when it exhibits *consistency*, *i.e.*: the MLE gets closer to the true parameter vector θ as more data is collected and included in the likelihood (assuming the model and its parameters remain constant during the data collection). More precise, the estimator $\hat{\theta}_n$ is consistent if

$$\Pr(|\hat{\theta}_n - \theta|) \to 0 \qquad \text{as} \qquad n \to \infty.$$

3.1 Consistency

In this section we prove that the MLE based on the Galerkin approximation $p_k(x,y,t)$ to the governing Fokker-Planck equation (4) is consistent. To do so, we first show that the distribution on Z based on the approximate $p_k(x,y,t)$ converges to the true distribution F:

$$\lim_{k\to\infty} F_k(z|\theta) = F(z|\theta).$$

We define $F_k(z|\theta)$ as

$$F_k(z|\theta) = \int_{-\infty}^{x}\int_{-\infty}^{y}\int_{-\infty}^{a_x}\int_{-\infty}^{a_y}\int_{-\infty}^{b_x}\int_{-\infty}^{b_y} \frac{\partial^4}{\partial a_x' \partial b_x' \partial a_y' \partial b_y'} q_k(x',y',t)\; dx'\,dy'\,da_x'\,db_x'\,da_y'\,db_y', \tag{15}$$

$$:= \Pr_k\left(X(t)\le x, Y(t)\le y, \min_{t'\le t} X(t') \le a_x, \max_{t'\le t} X(t') \le b_x, \min_{t'\le t} Y(t') \le a_y, \max_{t'\le t} Y(t') \le b_y\right) \tag{16}$$

where q_k is the approximation to q based on the p_k. Put simply, we will show that the distribution based on the approximate density approaches the true density as $k\to\infty$.

A key idea in the proof will be the interpretation of the measure $q(x,y,t)$

Lemma 1. *For F_k and F defined above,* $\lim_{k\to\infty} F_k(z) = F(z)$.

Proof. **do not read below; I'm working on it tonight** The weak solution $q_k(x,y,t)$ has the property that for any $C^\infty(\Omega)$ function $v \in L^2(\Omega)$,

$$\lim_{k\to\infty}(q_k, v) = (q, v),$$

where $(v,u) := \int_{a_x}^{b_x}\int_{a_y}^{b_y} v(x,y,t)u(x,y,t)dxdy$. and q is the strong solution.

Recalling equation (3), the strong solution has the probabilistic interpretation

$$\begin{aligned} q(x,y,t) &= \Pr\left(X(t)\in dx, Y(t)\in dy, a_x < X(t') < b_x, a_y < Y(t') < b_y | 0 < t' \le t, X(0)=x_0, Y(0)=y_0, \theta\right) \\ &= \Pr\left(X(t)\in dx, Y(t)\in dy, \min_{t'} X(t') \ge a_x, \max_{t'} X(t') \le b_x, \min_{t'} Y(t') \ge a_y, \max_{t'} Y(t') \le b_y \,\middle|\, 0 < t' \le t, X(0)=x_0, Y(0)=y_0, \theta\right) \end{aligned}$$

□

Lemma 2. *The maximum likelihood estimator is consistent as $n \to \infty$ and $k \to \infty$:*

$$\hat{\theta}_{n,k} \to \theta$$

.

Proof. **It's a bit sloppy. Also, we do not need asymptotic efficiency to prove equation (17)** By Lemma 1

$$X_k \xrightarrow{d} X \text{ as } k \to \infty.$$

Next, given Theorem 4.1 in Singler [2008], we know that, for each k, q_k is analytic in both the diffusion parameters and boundary parameters. Hence, the probability density function satisfies the criteria A1 - A6 in Casella and Berger [2002] to guarantee that, for data $X_k \sim F_k(\theta)$,

$$\hat{\theta}_{n,k}(X_k) \xrightarrow{p} \theta \text{ as } n \to \infty.$$

Now we need to show that the same holds for data sampled from F as $k \to \infty$. To do this, we will use Chebyshev's inequality:

$$\Pr_X\left(\left|\hat{\theta}_{n,k}(X) - \theta\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}_X\left[(\hat{\theta}_{n,k}(X) - \theta)^2\right]}{\varepsilon^2}.$$

By the Maximum theorem [REFERENCE], $\hat{\theta}_{n,k}(x)$ is a continuous function with respect to x, and further because we have bounded $\hat{\theta}$ from below and above,

$$\mathrm{E}_{X_k}\left[(\hat{\theta}_{n,k}(X_k) - \theta)^2\right] \to \mathrm{E}_X\left[(\hat{\theta}_{n,k}(X) - \theta)^2\right] \text{ as } k \to \infty$$

by the portmanteau lemma. Finally, we can show that

$$\mathrm{E}_{X_k}\left[(\hat{\theta}_{n,k}(X_k) - \theta)^2\right] \to 0 \text{ as } n \to \infty, \tag{17}$$

since the expected value of the estimator tends to θ and its variance goes to 0 when $n \to \infty$. Therefore, given any $\varepsilon > 0$ and $\delta > 0$, we can find a sufficiently large n and k such that

$$\Pr_X\left(\left|\hat{\theta}_{n,k}(X) - \theta\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}_X\left[(\hat{\theta}_{n,k}(X) - \theta)^2\right]}{\varepsilon^2} < \delta$$

□

References

JH Bramble, AH Schatz, V Thomée, and LB Wahlbin. Some convergence estimates for semidiscrete galerkin type approximations for parabolic equations. *SIAM Journal on Numerical Analysis*, 14(2):218–241, 1977.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Wai-Ki Ching, Jia-Wen Gu, and Harry Zheng. On correlated defaults and incomplete information. *arXiv preprint arXiv:1409.1393*, 2014.

Helen Haworth, Christoph Reisinger, and William Shaw. Modelling bonds and credit default swaps using a structural model with contagion. *Quantitative Finance*, 8(7):669–680, 2008.

Hua He, William P Keirstead, and Joachim Rebholz. Double lookbacks. *Mathematical Finance*, 8(3):201–228, 1998.

Enrique Ter Horst, Abel Rodriguez, Henryk Gzyl, and German Molina. Stochastic volatility models including open, close, high and low prices. *Quantitative Finance*, 12(2):199–212, 2012.

Steven Kou, Haowen Zhong, et al. First-passage times of two-dimensional brownian motion. *Advances in Applied Probability*, 48(4):1045–1060, 2016.

Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.

Douglas H Norrie and GJ De Vries. *The finite element method: fundamentals and applications*. Academic Press, 1973.

Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer-Verlag, 1989.

L Christopher G Rogers and Stephen E Satchell. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, pages 504–512, 1991.

Laura Sacerdote, Massimiliano Tamborrino, and Cristina Zucca. First passage times of two-dimensional correlated processes: Analytical results for the wiener process and a numerical method for diffusion processes. *Journal of Computational and Applied Mathematics*, 296:275–292, 2016.

John R Singler. Differentiability with respect to parameters of weak solutions of linear parabolic equations. *Mathematical and Computer Modelling*, 47(3):422–430, 2008.

Ivar Stakgold and Michael J Holst. *Green's functions and boundary value problems*, volume 99. John Wiley & Sons, 2011.

Eberhard Zeidler. Applied functional analysis: main principles and their applications. *Applied mathematical sciences*, 109, 1995.