Title

Volatility Estimation Methods for High-Frequency and Bivariate Open, Close, High, Low Prices

Author

Dinolov, Georgi Svetoslavov

Publication Date

2019

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

VOLATILITY ESTIMATION METHODS FOR HIGH-FREQUENCY AND BIVARIATE OPEN, CLOSE, HIGH, LOW PRICES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS AND APPLIED MATHEMATICS

by

Georgi Dinolov

March 2019

The Dissertation of Georgi Dinolov
is approved:

Abel Rodríguez, Chair

Hongyun Wang

Raquel Prado

Daniele Venturi

Lori Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Volatility Estimation Methods for High-Frequency and Bivariate Open, Close, High, Low Prices

by

Georgi Dinolov

Statistical models of price volatility most commonly use low-frequency (daily, weekly, or monthly) returns. However, despite their availability, two types of financial data have not been extensively studied: high-frequency data where sampling periods are on the order of seconds; and open, close, high, and low (OCHL) data which incorporate intraperiod extremes.

The first part of this dissertation focuses on the development of a filtering-based method for the estimation of volatility in high-frequency returns, which contrasts currently popular averaging-based approaches. The second part of this dissertation develops a foundational and novel method for likelihood-based estimation for bivariate OCHL, an approach unfeasible until now.

In Chapter 2, we formulate a discrete-time Bayesian stochastic volatility model for high-frequency stock-market data that directly accounts for microstructure noise, and outline a Markov chain Monte Carlo algorithm for parameter estimation. The methods described in this paper are designed to be coherent across all sampling timescales, with the goal of estimating the latent log-volatility signal from data collected at arbitrarily short sampling periods. In keeping with this goal, we carefully develop a method for eliciting priors. The empirical results derived from both simulated and real data show that directly accounting for microstructure noise in a state-space formulation allows for well-calibrated estimates of the log-volatility process driving prices.

In Chapter 3, we present and motivate the bivariate OCHL problem, enumerate the fundamental limitations of some common out-of-the-box approaches, and present a semidiscrete Galerkin numerical solver for computing the likelihood of the observed data. In addition, we prove the consistency of maximum likelihood estimates under the approximate density given by the solver.

Chapter 4 develops a closed-form, analytic solution to the OCHL likelihood problem in parameter ranges where the Galerkin solver requires near-infinite compute time and memory to produce numerically accurate results. A matching solution is also proposed to interpolate between parameter regions where neither the Galerkin nor analytic solutions are applicable. Thus, we present a method for producing likelihoods based on OCHL data over all model parameter ranges, which is a key requirement in statistical estimation algorithms. We use numerical experiments in both Chapters 3 and 4 to show the validity of our methods and demonstrate the increase in statistical power in estimating price volatility and correlation when using bivariate OCHL data.

In loving dedication to my wife, Alison, who makes it all worth it.

Acknowledgments

I want to thank my advisers, Professor Abel Rodríguez and Professor Hongyun Wang, for helping and mentoring me through completing this work. Thank you both for your unyielding patience and never losing faith in me. I have learned tremendously, and I cannot have done this without you both guiding me at every step. I also want to thank both Professor Raquel Prado and Professor Athanasios Kottas for advising me during my first as a statistics graduate student at UC Santa Cruz, and Professor Prado in particular for serving on my dissertation committee. Finally, I want to thanks Professor Daniele Venturi for also serving on my committee and providing an invaluable perspective.

I want to thank my wife, Alison, for always supporting me and reminding me of the light at the end of the tunnel. Thank you also to the rest of my family for your encouragement and love.

Chapter 1

Introduction

Financial markets have been the subject of research since at least Bachelier [1900], and serious efforts to characterize price variability continued through the 20^{th} century, with a more modern resurgence of interest initiated by Mandelbrot [1967]. Although long in use by market participants in the form of heuristic rules prior to 1973 [Haug and Taleb, 2011], the Black-Scholes-Merton formula [Black and Scholes, 1973, Merton, 1973] allowed for statistical models of price volatility to be used in option pricing and trading decisions. This in turn intensified the need to quantify price dynamics and more deeply study their statistical properties. Engle [1982] and Bollerslev [1986] introduced the Autoregressive Conditionally Heteroscedastic (ARCH) and Generalized Autoregressive Conditionally Heteroscedastic (GARCH) models as state-space-like time series representation of price volatility. Shortly thereafter followed the stochastic volatility model of Hull and White [1987], which is still popular today. Extensions and further applications of these models to different assets types, markets, pricing scenarios, and inferential settings have allowed the literature to grow tremendously. In addition to the strides made by researchers in the field, the wide availability of high-quality data has facilitated the study of price volatility from a statistical standpoint. Despite their availability, two particular types of data have not been as extensively studied in the literature: high-frequency time series data and open, close, high, and low (OCHL for short) data. This work develops statistical approaches that more efficiently use these types of data.

1.1 High-Frequency Data

Traditionally, financial models have used low-frequency returns (e.g., daily, weekly or monthly returns) to investigate price volatility. However, as high-frequency price data (which is roughly defined as sampling every 5 minutes or less) has become widely available, interest has turned to generating high-resolution estimates of the volatility path and to improving estimates of volatility model parameters over higher frequencies.

Attempts have been made to fit the GARCH model to high-frequency data [Bollerslev, 1986, Andersen and Bollerslev, 1997b]. However, the assumptions behind the classical GARCH model are not robust with respect to the specification of the sampling interval, and therefore the model is not invariant to temporal aggregation [Drost and Nijman, 1993, Andersen and Bollerslev, 1997b, Zumbach, 2000]. To address this issue some authors have turned to estimating low-frequency GARCH and stochastic volatility models using relevant summaries of the high-frequency prices. One popular summary statistic is the realized variance estimator [Comte and Renault, 1998, Andersen et al., 2001b, Barndorff-Nielsen and Shephard, 2002a]. The realized variance is defined as the sum of squared high-frequency log returns over the period of interest. Under strict stationarity and some other weak regularity conditions for the volatility process, the realized variance converges in probability to the integrated variance of the true diffusion process as the sampling frequency increases. Takahashi et al. [2009] and Shirota et al. [2014] use both high-frequency returns as well as the realized variance to estimate stochastic volatility models, while Hansen et al. [2012] do the same for GARCH models. Similarly, Maneesoonthorn et al. [2014] use the realized volatility and the bipower variation estimators to estimate stochastic volatility models with jumps, while Bollerslev and Zhou [2002] use high order powers of the realized volatility as approximations to higher orders of integrated volatility.

A key challenge in working with high frequency prices is that they are often contaminated with microstructure noise. Indeed, as the sampling period shrinks down to the transaction-by-transaction frequency, irregular spacing between transactions, discreteness in transaction prices, and very short term temporal dependence become dominant features of the data [Stoll, 2000]. One consequence of the presence of microstructure noise is that the realized variance becomes a biased and inconsistent estimator of the true integrated variance [Zhou, 1996]. Possible solutions to this issue have been proposed by Zhang et al. [2005], who suggest sampling data sparsely at an optimally determined frequency and then averaging across the possible grids over the data, Ait-Sahalia et al. [2011], who propose combining estimators based on subsampling data at different frequencies, and Hansen and Lunde [2006], Barndorff-Nielsen et al. [2008], who employ a class of kernel-based methods similar to those used for estimating the long-run variance of a stationary time-series in the presence of autocorrelation. In the context of model-based approaches it is common to assume that the summaries of the high-frequency returns used to estimate the model are noisy versions of the true realized volatilities (e.g., see Venter and de Jongh [2012], Shirota et al. [2014]).

Chapter 1 describes a Bayesian stochastic volatility model for high-frequency data that explicitly accounts for the presence of microstructure noise. Unlike other approaches in the literature, we estimate our model directly using the high-frequency price data rather than summaries of the high-frequency returns. To account for the effect of microstructure noise we introduce a hierarchical specification in which the observed high-frequency prices are noisy versions of the true unknown prices. One appealing feature of our proposed model is that it is (approximately) coherent across all sampling frequencies, which is in line with previous efforts to validate the application of discrete-time models for volatility in high-frequency settings [Andersen et al., 1999]. Coherency is achieved by starting with a continuous-time model and then carefully discretizing the exact solution to the stochastic differential equations for the price and volatility processes, and by carefully eliciting prior distributions for the parameters of the continuous-time model.

1.2 Bivariate OCHL Data

Chapter 1 is concerned with high-frequency data where only the period opening and closing prices are used to estimate volatility. However, even for lower-frequency data, it is common to assume that asset prices follow correlated geometric Brownian motions. In that setting, estimates of volatility and correlation are usually based on log-returns, which depend only on the opening and closing prices. However, typically more information is available. For example, maximum and minimum trading prices over different trading intervals are readily available which can serve as summary statistics for volatility over the periods of interest. In univariate settings, it has been shown that incorporating this additional information can substantially improve estimates. For example, Horst et al. [2012] derive a full likelihood-based (Bayesian) approach to estimate volatility in univariate financial time series 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. Likelihood-free approaches, like that of Rogers and Satchell [1991], Rogers et al. [2008], on the other hand suffer from not being able to be easily integrated into inferential frameworks that require explicit estimates of probability, as well as being sub-optimal. Likelihood approaches involving less than all four extrema have 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], and to pricing financial derivative instruments whose payoff depends on some (but not all) of the observed boundaries [He et al., 1998].

A general likelihood-based approach for incorporating OCHL data in bivariate model estimation has, until now, been unfeasible. This is due to the lack of a robust and numerically efficient solution, necessary when requiring repeated computation, to the partial differential equation determining this likelihood. In Chapters 2 and 3 we develop this numerical solution and demonstrate the increase in statistical power in using OCHL data when comparing to existing methods.

Chapter 2

High-Frequency Stochastic Volatility

2.1 Introduction

Estimating asset price volatilities is a common problem in finance; for example, accurate estimates of volatility paths play a key role in both option pricing and portfolio design. Traditionally, financial models have used low-frequency returns (e.g., daily, weekly or monthly returns) to investigate price volatility. Early attempts at incorporating higher-frequency information focused on using intra-period maximum and minimum prices (e.g., see Alizadeh et al. [2002b], Brandt and Diebold [2003], and Chou et al. [2010]). However, as high-frequency price data has become widely available, interest has turned to using all intra-period prices to generate high-resolution estimates of the volatility path and to improve estimates of the integrated volatility over higher frequencies. In this chapter we describe a Bayesian stochastic volatility model for high-frequency data that explicitly accounts for the confounding effect of microstructure noise that normally makes difficult the direct application of such dynamical models to this type of data. In this way, we avoid loss of informativeness that is present when using summaries of the data, such as the realized volatility estimators, instead of fitting to the data directly.

The remainder of the Chapter is structured as follows: Section 2.2 describes the continuous- and discrete-time version of the model. Section 2.3 details the priors used and the method through which they were derived. Section 2.4 outlines the Bayesian Markov chain Monte Carlo (MCMC) algorithm used to fit the model. Section 2.5 examines the effect of certain model parameters on the posterior variance of the mean volatility level in our model. Finally, Section 2.6 includes simulation results demonstrating the robustness of our inferential procedure to microstructure noise.

2.2 Model Formulation

We begin with the continuous-time stochastic volatility model of Hull and White [1987], where the price $\hat{S}_t$ of an asset follows a Geometric Brownian motion and the time-varying log-volatility process $\log(\hat{\sigma}_t)$ follows a mean-reverting Ornstein-Uhlenbeck (OU) process,

$$d\log(\hat{S}_t) = \hat{\mu}\, dt + \hat{\sigma}_t \sqrt{dt}\hat{\epsilon}_t, \tag{2.1}$$

$$d\log(\hat{\sigma}_t) = -\hat{\theta}(\log(\hat{\sigma}_t) - \hat{\alpha})\, dt + \hat{\tau} \sqrt{dt}\hat{\epsilon}_{t,1}, \tag{2.2}$$

where $\hat{\epsilon}_t$ and $\hat{\epsilon}_{t,1}$ are dependent Wiener processes with instantaneous correlation ρ. This model not only allows for the volatility to evolve over time, but also captures leverage effects though the correlation between $\hat{\epsilon}_{t,1}$ and $\hat{\epsilon}_{t,2}$ (e.g., see Black [1976]). However, since at least the papers by Barndorff-Nielsen and Shephard [2001] and

Chernov et al. [2003], it is understood that single-factor volatility models cannot reproduce the dependence structure of asset returns. Further, at high sampling frequencies, jumps have been shown to be an important part of asset price variance [Huang and Tauchen, 2005]. As a consequence of these results, we extend the model in (2.1)-(2.2) to include two factors for the volatility and jumps in the returns:

$$d\log(\hat{S}_t) = \hat{\mu}\, dt + \sqrt{\hat{\sigma}_{t,1}\hat{\sigma}_{t,2}}\,\sqrt{dt}\hat{\epsilon}_{t,1} + dJ_t, \tag{2.3}$$

$$d\log(\hat{\sigma}_{t,1}) = -\hat{\theta}_1(\log(\hat{\sigma}_{t,1}) - \hat{\alpha})\, dt + \hat{\tau}_1\,\sqrt{dt}\hat{\epsilon}_{t,1}, \tag{2.4}$$

$$d\log(\hat{\sigma}_{t,2}) = -\hat{\theta}_2(\log(\hat{\sigma}_{t,2}) - \hat{\alpha})\, dt + \hat{\tau}_2\,\sqrt{dt}\hat{\epsilon}_{t,2}. \tag{2.5}$$

The volatility exhibited by the log-prices in (2.3) is specified as a product of two OU processes. For the purposes of this paper, we will think of the volatility processes as distinct, $\hat{\sigma}_{t,1}$ being slow and $\hat{\sigma}_{t,2}$ being fast. By slow versus fast we mean that the mean-reversion timescale of (2.4) is greater than that of (2.5): $1/\hat{\theta}_1 \gg 1/\hat{\theta}_2$. We allow for leverage between the innovations of returns and the fast volatility, but the slow volatility and the price innovations are independent; the two volatility processes are also independent:

$$\mathrm{E}\left[\hat{\epsilon}_t\hat{\epsilon}_{t,2}\right] = \rho, \qquad \mathrm{E}\left[\hat{\epsilon}_t\hat{\epsilon}_{t,1}\right] = 0, \qquad \mathrm{E}\left[\hat{\epsilon}_{t,1}\hat{\epsilon}_{t,2}\right] = 0.$$

Finally, we model the jump process dJ_t as a compound Poisson process with constant jump intensity λ and i.i.d. jump size $Z_t \sim N(\mu_J, \sigma_J^2)$. In other words, over a finite interval Δ, $J(t+\Delta) - J(t) = \sum_{j=1}^{N(\Delta)} Z_{t_j}$, where

$$(t_j - t_{j-1}) \sim \mathrm{Exp}\,(\lambda)\,, \quad N(\Delta) \sim \mathrm{Pois}\,(\lambda\Delta)\,, \quad Z_{t_j} \sim N(\mu_J, \sigma_J^2), \quad t_j \in (t, t+\Delta). \tag{2.6}$$

We assume that jump sizes and arrival times are independent of either the price or volatility processes.

To generate a discretization of the model in (2.3) - (2.5), consider the (strong) solution of the Ornstein-Uhlenbeck process governing the evolution of the log-volatility in (2.2),

$$\log(\hat{\sigma}_t) \sim N\left(\hat{\alpha} + \exp\left\{-\hat{\theta}t\right\}\left\{\log(\hat{\sigma}_0) - \hat{\alpha}\right\}, \frac{\hat{\tau}^2}{2\hat{\theta}}\left\{1 - \exp(-2\hat{\theta}t)\right\}\right), \tag{2.7}$$

with stationary distribution

$$\log(\hat{\sigma}_t) \sim N\left(\hat{\alpha}, \frac{\hat{\tau}^2}{2\hat{\theta}}\right). \tag{2.8}$$

For an arbitrary time interval Δ (which, for the purpose of this paper, we measure in milliseconds) we can use (2.7) to generate the finite-difference equations

$$\log(S_j) = \log(S_{j-1}) + \mu(\Delta) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\,\epsilon_j + J_j(\Delta), \tag{2.9}$$

$$\log(\sigma_{j+1,1}) = \alpha(\Delta) + \theta_1(\Delta)\left\{\log(\sigma_{j,1}) - \alpha(\Delta)\right\} + \tau_1(\Delta)\,\epsilon_{j,1}, \tag{2.10}$$

$$\log(\sigma_{j+1,2}) = \alpha(\Delta) + \theta_2(\Delta)\left\{\log(\sigma_{j,2}) - \alpha(\Delta)\right\} + \tau_2(\Delta)\,\epsilon_{j,2}, \tag{2.11}$$

where $j = 0, 1, \ldots, \lfloor T/\Delta \rfloor$, $i = 1, 2$, and

$$\sigma_{j+1,i} = \hat{\sigma}_{(j+1)\Delta,i}\sqrt{\Delta}, \qquad S_j = \hat{S}_{j\Delta}, \qquad J_j(\Delta) = J((j+1)\Delta) - J(j\Delta) \tag{2.12}$$

$$\begin{gathered} \alpha(\Delta) = \hat{\alpha} + \frac{1}{2}\log(\Delta), \quad \mu(\Delta) = \hat{\mu}\Delta, \\ \tau_i(\Delta) = \hat{\tau}_i\sqrt{\frac{1 - \exp\left\{-2\hat{\theta}_i\Delta\right\}}{2\hat{\theta}_i}}, \quad \theta_i(\Delta) = \exp\left\{-\hat{\theta}_i\Delta\right\}, \end{gathered} \tag{2.13}$$

and

$$\begin{pmatrix} \epsilon_j \\ \epsilon_{j,1} \\ \epsilon_{j,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix} \right).$$

We write $\alpha(\Delta)$, $\mu(\Delta)$, $\theta_i(\Delta)$, and $\tau_i(\Delta)$ to emphasize that we have a different set of parameters depending on the choice of Δ. The random variable $J_j(\Delta)$ is defined in (2.6) and is the jump increment from $j\Delta$ to $(j+1)\Delta$. Finally, (2.9) is an approximate discretization, while (2.10) and (2.11) are exact solutions.

Using the strong solution in (2.7) to derive the finite difference equations (2.9) - (2.11) allows us to take any step size Δ irrespective of the relative magnitude of the continuous-time model parameters. Indeed, the more standard forward-Euler discretization provides a poor approximation to the continuous-time model when $\Delta > 1/\hat{\theta}$, the timescale of inertia of the log-volatility process. Even in cases where $\Delta < 1/\hat{\theta}$, discretizing the model SDE using the exact solution of the OU process is a more accurate approximation of the finite-time transition density of the continuous process in (2.3) - (2.5) [Elerian et al., 2001].

The discretization (2.9) - (2.11) is a discrete-time approximation of the stochastic process in (2.3) - (2.5) and the likelihood based on this discretization introduces a bias when estimating continuous time parameters. Elerian et al. [2001] and Eraker [2001] both show that a way to correct for this bias (in the context of MCMC analysis) is to introduce latent, intra-interval sub-samples of the process. We expect to achieve a similar, but better, result by decreasing Δ and accounting for microstructure noise. Moreover, the inverse of the transformations in (2.13) make it possible to meaningfully compare parameters inferred from different sampling frequencies and thereby check the coherency of our inferential procedure across different timescales.

In order to account for the effect of microstructure noise, we extend the previous model by differentiating between the *true* log asset price $\log(S_j)$ and the *discretely observed* log price $Y_j = \log(P_j)$. We treat these observed log prices as a noise-contaminated version of the true log price which is observed discretely only $n(\Delta) = \lfloor T/\Delta \rfloor$ times and whose index j corresponds to $j\Delta$ in continuous-time. More specifically, we let

$$Y_j = \log(S_j) + \zeta_j, \tag{2.14}$$

where $\zeta_1, \zeta_2, \ldots$ are independent and identically distributed errors with mean zero and standard deviation ξ. To motivate (2.14), consider one possible source of microstructure noise, the bid-ask spread. We can think of the idealized, "true" equilibrium price as evolving continuously in time by being moved by market supply and demand. Real-time order arrival and market friction makes it so that transaction prices are recorded at the highest bid or lowest ask levels only, thereby bounding the equilibrium market price in the bid-ask range. In this case it is natural to assume that $P_j = S_j + \nu_j$, where $\nu_j \sim U[-D_p/2, D_p/2]$ and D_p is the size of the bid-ask spread. Using a first-order Taylor approximation then leads to $Y_j = \log(P_j) \approx \log(S_j) + \zeta_j$, where $\zeta_j = \frac{1}{S_j}\nu_j$. A similar argument can be used to account for the effect of price discretization.

More generally, in order to account for the approximation error as well as for other sources of microstructure noise, we let $\zeta_t \sim N(0, \xi^2)$ where $\xi \approx \frac{D}{2Q}$, $D = \max\{D_p, D_s\}$, D_p represents a rough estimate of the bid-ask spread over the period of interest, D_s represents another possible source of microstructure noise (such as price quantized to the nearest cent), and Q is a rough guess of the average price of the asset over the period of interest. Note that the distribution of ξ is independent of the time scale Δ used for the discretization of the continuous-time process, and therefore independent of the frequency at which prices are observed.

To summarize, our hierarchical discrete-time stochastic volatility model reduces to

$$\begin{aligned}
Y_j &= \log(S_j) + \zeta_j, &(2.15)\\
\log(S_j) &= \log(S_{j-1}) + \mu(\Delta) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\,\epsilon_j + J_j(\Delta), &(2.16)\\
\log(\sigma_{j+1,1}) &= \alpha(\Delta) + \theta_1(\Delta)\left\{\log(\sigma_{j,1}) - \alpha(\Delta)\right\} + \tau_1(\Delta)\,\epsilon_{j,1}, &(2.17)\\
\log(\sigma_{j+1,2}) &= \alpha(\Delta) + \theta_2(\Delta)\left\{\log(\sigma_{j,2}) - \alpha(\Delta)\right\} + \tau_2(\Delta)\,\epsilon_{j,2}, &(2.18)
\end{aligned}$$

where

$$\zeta_j \sim N(0, \xi^2), \qquad \begin{pmatrix} \epsilon_j \\ \epsilon_{j,1} \\ \epsilon_{j,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix} \right),$$

and initial conditions

$$\log(\sigma_{0,i}) \sim N\left(\alpha, \frac{\tau_i(\Delta)^2}{1-\theta_i(\Delta)^2}\right), \qquad \log(S_0) \sim N\left(\eta, \kappa^2\right),$$

where the priors for $\log(\sigma_{0,i})$ are the marginal stationary distributions of the discrete-time, autoregressive volatility processes. Table 2.1 summarizes our notation.

2.3 Prior Elicitation

We approach the problem of estimation and prediction for the model described above using Bayesian methods. This requires that we elicit priors for the unknown

	Symbol	Interpretation
Data	P_j	Observed asset price at time $j\Delta$.
	Y_j	Logarithm of the observed asset price at time $j\Delta$.
Parameters	S_j	True asset price at time $j\Delta$.
	$\sigma_{j,i}$	Discrete-time approximation of the slow/fast volatility i of the true asset price at time $j\Delta$.
	$\alpha(\Delta)$	Discrete-time approximation of the stationary mean of the log volatility.
	$\mu(\Delta)$	Discrete-time approximation to the mean asset return.
	$\theta_i(\Delta)$	Discrete-time approximation to the autocorrelation associated with the log volatility i.
	$\tau_i(\Delta)$	Discrete-time approximation to the volatility of volatility i.
	ρ	Correlation coefficient between volatility and price innovations for the slow volatility process.
	ξ	Standard deviation associated with the microstructure noise.
	λ	Arrival rate of price jumps.
	μ_J	Expected value of price jump sizes.
	σ_J	Standard deviation of price jump sizes.
Hyper	Δ	Time step between observations (Fixed).
	η	Mean for the true asset price at time 0. (Fixed; no inference performed)
	κ	Standard deviation for the true asset at time 0. (Fixed; no inference performed)

Table 2.1: Notation summary for our high-frequency stochastic volatility model, including data, parameters and hyper-parameters.

parameters ρ, ξ^2, $\alpha(\Delta)$, $\mu(\Delta)$, $\theta(\Delta)$ and $\tau(\Delta)$. Eliciting a prior for the correlation parameter ρ and the microstructure variance ξ^2 is relatively straightforward since their value and interpretation are independent of the time step Δ. On the other hand, ensuring that the priors for $\alpha(\Delta)$, $\mu(\Delta)$, $\theta(\Delta)$ and $\tau(\Delta)$ are coherent across scales, i.e., that the priors provide the same information no matter what the time step Δ is, is non trivial. To address this problem we proceed to elicit priors on the continuous-time parameters $\hat{\alpha}$, $\hat{\mu}$, $\hat{\theta}$, $\hat{\tau}$ and then use the formulas in (2.13) to obtain the implied priors on $\alpha(\Delta)$, $\mu(\Delta)$, $\theta(\Delta)$ and $\tau(\Delta)$ for any time step Δ. Ideally, such priors would be invariant to the transformations in (2.13). However, fully invariant priors are difficult to elicit and would, in any case, lead to computationally complicated models even using simulation-based methods such as Markov chain Monte Carlo algorithms. Hence, we settle for the more modest goal of assigning priors that belong to families that are conditionally conjugate and therefore lead to computationally tractable models, but whose first two moments are (approximately) coherent across scales.

1. **Prior for $\boldsymbol{\rho}$**: We assign $(\rho+1)/2$ a symmetric beta distribution with mean 1/2 and precision c. This prior ensures that $\rho \in [-1, 1]$ as required and implies that $E(\rho) = 0$ a priori. Furthermore, for large values of c, this means that we believe a priori that the leverage effect is relatively small.

2. **Prior for $\boldsymbol{\mu}(\boldsymbol{\Delta})$**: For the mean of the asset returns a prior in the normal family leads to a simple full conditional distribution for MCMC sampling. If we let $\mathrm{E}[\hat{\mu}] = \hat{a}_{\hat{\mu}}$ and $\mathrm{Var}[\hat{\mu}] = \hat{b}^2_{\hat{\mu}}$, then $\mu(\Delta) = \hat{\mu}\Delta$ leads to $\mathrm{E}[\mu(\Delta)] = \Delta\hat{a}_{\hat{\mu}}$ and $\mathrm{Var}[\mu(\Delta)] = \Delta^2\hat{b}^2_{\hat{\mu}}$. Hence, in our analysis we use the prior

$$\mu(\Delta) \sim N(\Delta\hat{a}_{\hat{\mu}}, \Delta^2\hat{b}^2_{\hat{\mu}})$$

   where values of $\hat{a}_{\hat{\mu}}$ and $\hat{b}^2_{\hat{\mu}}$ are elicited from historical data.

3. **Prior for $\boldsymbol{\theta}_i(\boldsymbol{\Delta})$**: The discrete-time autocorrelation coefficient $\theta(\Delta)$ of the volatility process is bounded above by 1 and below by 0 such that $\log(\sigma_j)$ is bounded as $j \to \infty$. Hence, we employ a truncated normal prior for $\theta(\Delta)$,

$$p(\theta(\Delta)) \propto N\left(a_\theta(\Delta), b^2_\theta(\Delta)\right) \mathbb{1}_{(\theta(\Delta)\in[0,1])},$$

   which leads again to a tractable computational algorithm. Note that because of the truncation,

$$\mathrm{E}\left[\theta(\Delta)\right] = a_\theta(\Delta) + \frac{\phi\left(-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\right) - \phi\left(\frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\right)}{\Phi\left(\frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\right) - \Phi\left(-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\right)} b_\theta(\Delta) \tag{2.19}$$

$$\mathrm{Var}\left[\theta(\Delta)\right] = b_\theta^2(\Delta)\left[1 + \frac{-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\phi\left(-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\right) - \frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\phi\left(\frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\right)}{\Phi\left(\frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\right) - \Phi\left(-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\right)} + \left\{\frac{\phi\left(-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\right) - \phi\left(\frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\right)}{\Phi\left(\frac{1-a_\theta(\Delta)}{b_\theta(\Delta)}\right) - \Phi\left(-\frac{a_\theta(\Delta)}{b_\theta(\Delta)}\right)}\right\}^2\right] \tag{2.20}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution functions of the standard normal distribution. Now, given the prior mean $\hat{a}_{\hat{\theta}}$ and variance $\hat{b}_{\hat{\theta}}^2$ for $\hat{\theta}$, we choose the values of $a_\theta(\Delta)$ and $b_\theta(\Delta)$ so that the mean and variance of $\theta(\Delta)$ above are approximately equal to the mean and variance of $\exp\left\{-\hat{\theta}\Delta\right\}$. To simplify calculation of the moments of $\exp\left\{-\hat{\theta}\Delta\right\}$ we use a second-order Taylor expansion of $\exp\left\{-\hat{\theta}\Delta\right\}$ to approximate the first two moments of $\theta(\Delta)$ in terms of $\hat{a}_{\hat{\theta}}$ and $\hat{b}_{\hat{\theta}}^2$, an approach known as the Delta-Method (e.g., see Casella and Berger [2002]):

$$\mathrm{E}\left[\exp\left\{-\hat{\theta}\Delta\right\}\right] \approx \exp\left(-\hat{a}_{\hat{\theta}}\Delta\right)\left(1 + \frac{1}{2}\hat{b}_{\hat{\theta}}^2\Delta^2\right), \tag{2.21}$$

$$\mathrm{E}\left[\exp\left\{-2\hat{\theta}\Delta\right\}\right] \approx \exp\left(-2\hat{a}_{\hat{\theta}}\Delta\right)\left(1 + 2\hat{b}_{\hat{\theta}}^2\Delta^2\right). \tag{2.22}$$

Using (2.19), (2.20), (2.21), and (2.22), and by setting

$$\mathrm{E}\left[\theta(\Delta)\right] = \mathrm{E}\left[\exp\left\{-\hat{\theta}\Delta\right\}\right], \qquad \mathrm{Var}\left[\theta(\Delta)\right] = \mathrm{Var}\left[\exp\left\{-\hat{\theta}\Delta\right\}\right],$$

we obtain a system of two equations with two unknowns that can be solved numerically to find the values of $a_\theta(\Delta)$ and $b_\theta^2(\Delta)$ in terms of $\hat{a}_{\hat{\theta}}$, $\hat{b}_{\hat{\theta}}^2$, and Δ. To elicit $\hat{a}_{\hat{\theta}}$ and $\hat{b}_{\hat{\theta}}^2$, recall that $\hat{\theta}$ is the inverse of the time scale of inertia for

$\log(\hat{\sigma}_t)$ in the continuous-time formulation, which can be thought of as the characteristic time length, or unit of time, over which the process for the diffusion of $\log(\hat{\sigma}_t)$ "forgets" about an endogenous shock. The two hyperparameters can be chosen so that the prior probability mass for $\hat{\theta}$ permits a reasonable range for the timescale of inertia.

4. **Prior for $\boldsymbol{\alpha}(\boldsymbol{\Delta})$**: For the mean log-volatility level $\hat{\alpha}$, we once again use a computationally convenient prior in the normal family. Letting $\mathrm{E}\left[\hat{\alpha}\right] = \hat{a}_{\hat{\alpha}}$ and $\mathrm{Var}\left[\hat{\alpha}\right] = \hat{b}^2_{\hat{\alpha}}$, and recalling that $\alpha(\Delta) = \hat{\alpha} + \frac{1}{2}\log(\Delta)$, we have

$$\alpha(\Delta) \sim N\left(\hat{a}_{\hat{\alpha}} + \frac{1}{2}\log(\Delta), \hat{b}^2_{\hat{\alpha}}\right).$$

   To elicit the values of $\hat{a}_{\hat{\alpha}}$ and $\hat{b}^2_{\hat{\alpha}}$, recall that $\hat{\alpha}$ is the stationary (long-term) median of the volatility process. Hence, for most assets these parameters could be elicited by looking at the time series of the asset's implied volatility (e.g., the VIX index if the asset is the S&P500 index).

5. **Prior for $\boldsymbol{\tau^2}(\boldsymbol{\Delta})$**: We use a prior in the Inverse-Gamma family for $\tau^2(\Delta)$, so that

$$\tau^2(\Delta) \sim \text{Inv-Gamma}\left(a_{\tau^2}(\Delta), b_{\tau^2}(\Delta)\right).$$

   To find the values of $a_{\tau^2}(\Delta)$ and $b_{\tau^2}(\Delta)$ recall that

$$\tau^2(\Delta) = \hat{\tau}^2\left(1 - \exp\{-2\hat{\theta}\Delta\}\right) / \left(2\hat{\theta}\right).$$

   If we let $\mathrm{E}\left[\hat{\tau}^2\right] = \hat{a}_{\hat{\tau}^2}$ and $\mathrm{Var}\left[\hat{\tau}^2\right] = \hat{b}^2_{\hat{\tau}^2}$, and if we use the prior mean and variance of $\hat{\theta}$ as before, we can again apply the Delta-Method to approximate the prior first and second moments of $\tau^2(\Delta)$ by performing a second-order Taylor expansion of $\tau^2(\Delta)$ and $(\tau^2(\Delta))^2$ about the prior means of $\hat{\tau}^2$ and $\hat{\theta}$, leading to

$$\begin{aligned}
\mathrm{E}\left[\hat{\tau}^2\left(\frac{1-\exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right] \approx{}& \hat{\tau}^2\left(\frac{1-\exp\left\{-2\hat{\theta}\Delta\right\}}{2\hat{\theta}}\right)\Bigg|_{\hat{\tau}^2=\hat{a}_{\hat{\tau}^2},\hat{\theta}=\hat{a}_{\hat{\theta}}} \\
&+\frac{1}{2}\hat{b}^2_{\hat{\tau}^2}\,\frac{\partial^2}{\partial^2\hat{\tau}^2}\left[\hat{\tau}^2\left(\frac{1-\exp\left\{-2\hat{\theta}\Delta\right\}}{2\hat{\theta}}\right)\right]\Bigg|_{\hat{\tau}^2=\hat{a}_{\hat{\tau}^2},\hat{\theta}=\hat{a}_{\hat{\theta}}} \\
&+\frac{1}{2}\hat{b}^2_{\hat{\theta}}\,\frac{\partial^2}{\partial^2\hat{\theta}}\left[\hat{\tau}^2\left(\frac{1-\exp\left\{-2\hat{\theta}\Delta\right\}}{2\hat{\theta}}\right)\right]\Bigg|_{\hat{\tau}^2=\hat{a}_{\hat{\tau}^2},\hat{\theta}=\hat{a}_{\hat{\theta}}}
\end{aligned} \tag{2.23}$$

and

$$\begin{aligned}
\mathrm{E}\left[\left\{\hat{\tau}^2\left(\frac{1-\exp\{-2\hat{\theta}\Delta\}}{2\hat{\theta}}\right)\right\}^2\right] \approx{}& \left\{\hat{\tau}^2\left(\frac{1-\exp\left\{-2\hat{\theta}\Delta\right\}}{2\hat{\theta}}\right)\right\}^2\Bigg|_{\hat{\tau}^2=\hat{a}_{\hat{\tau}^2},\hat{\theta}=\hat{a}_{\hat{\theta}}} \\
&+\frac{1}{2}\hat{b}^2_{\hat{\tau}^2}\,\frac{\partial^2}{\partial^2\hat{\tau}^2}\left\{\hat{\tau}^2\left(\frac{1-\exp\left\{-2\hat{\theta}\Delta\right\}}{2\hat{\theta}}\right)\right\}^2\Bigg|_{\hat{\tau}^2=\hat{a}_{\hat{\tau}^2},\hat{\theta}=\hat{a}_{\hat{\theta}}} \\
&+\frac{1}{2}\hat{b}^2_{\hat{\theta}}\,\frac{\partial^2}{\partial^2\hat{\theta}}\left\{\hat{\tau}^2\left(\frac{1-\exp\left\{-2\hat{\theta}\Delta\right\}}{2\hat{\theta}}\right)\right\}^2\Bigg|_{\hat{\tau}^2=\hat{a}_{\hat{\tau}^2},\hat{\theta}=\hat{a}_{\hat{\theta}}}.
\end{aligned} \tag{2.24}$$

The right sides of (2.23) and (2.24) are functions of $a_{\tau^2}(\Delta)$ and $b_{\tau^2}(\Delta)$, so that the above system of equations can be solved numerically to find $a_{\tau^2}(\Delta)$ and $b_{\tau^2}(\Delta)$ in terms of the other known prior hyper-parameters. Finally, to elicit $\hat{a}_{\hat{\tau}^2}$ and $\hat{b}_{\hat{\tau}^2}$,we have to recall that the ratio $\frac{\hat{\tau}^2}{2\hat{\theta}}$ represents the long-run variance of the log-volatility process $\log(\hat{\sigma})$. The prior for $\hat{\tau}^2$ can therefore be elicited from market-traded approximations of the volatility process, such as the VIX.

6. **Prior for $\boldsymbol{\xi^2}$**: For computational convenience, the variance of the microstructure noise is assigned an inverse Gamma with shape parameter a_ξ

and rate parameter b_ξ. The mean of the prior can be elicited from information about the bid-ask spread, the tick size and the average price of the stock as discussed in Section 2.2, while its standard deviation is selected so that we stay within an order of magnitude (above and below) of the mean.

7. **Prior for $(\boldsymbol{\lambda}, \boldsymbol{\mu_J}, \boldsymbol{\sigma_J^2})$:** We assign a Gamma$(a_\lambda, b_\lambda)$ prior for the arrival rate of jumps, and a Normal-Inverse-gamma prior for the expected value and variance of the jump sizes with hyper-parameters $(m_J, 1, a_J, b_J)$. The hyper-parameters for the arrival rate of jumps are elicited from weekly dynamics, assuming an average of once-per-week jump and a variance an order to magnitude greater:

$$\mathrm{E}\left[\lambda\Delta^*\right] = \Delta^* \frac{a_\lambda}{b_\lambda} = 1, \qquad \mathrm{Var}\left[\lambda\Delta^*\right] = (\Delta^*)^2 \frac{a_\lambda}{b_\lambda^2} = 10,$$

$$\Delta^* = 1 \text{ week (milliseconds)}.$$

   The jump size hyper-parameters are elicited from historical data of the S&P 500 index. In particular, m_J is the average observed weekend jump size from 1990 to 2015 on the S&P 500, whereas $b_J/(a_J - 1)$ is the observed variance of the said jumps. Finally, we set $a_J = 1.1$ so that the prior admits a valid expectation for the jump variance σ_J^2 but is minimally informative (equivalent to 1.1 data points in the prior).

2.4 Computation

The posterior distribution of high-frequency stochastic volatility is analytically intractable, so we perform parameter inference and prediction using a Markov chain Monte Carlo (MCMC) algorithm. Our sampler extends the ideas introduced in Omori et al. [2007], which used a mixture of Normals approximation to the distribution of a log χ^2 distribution. More specifically, our algorithm alternates between (1) integrating out the true latent prices with the forward component of the Forward-Backward algorithm [Carter and Kohn, 1994, Frühwirth-Schnatter, 1994] and sampling $(\rho, \xi, \mu(\Delta))$ jointly given all other parameters (2) sampling the true asset prices $S_0, S_1, \ldots, S_{n(\Delta)}$ from their joint full conditional distribution using the Forward-Backward algorithm, (3) jointly sampling the mixture indicators $\gamma_1, \ldots, \gamma_{n(\Delta)}$ (to be introduced below) given all other parameters, (4) given the mixture indicators, integrating out the volatilities using the forward component of the Forward-Backward algorithm and jointly sampling all of the volatility parameters except ρ, (5) jointly sampling the volatilities $(\sigma_{1,1}, \sigma_{1,2}), \ldots, (\sigma_{n(\Delta),1}, \sigma_{n(\Delta),2})$, $(\sigma_{n(\Delta)+1,1}, \sigma_{n(\Delta)+1,2})$ using a second Forward-Backward algorithm.

For Steps (1) and (2) in our inferential procedure, note that, given the mean return $\mu(\Delta)$, the microstructure variance ξ^2 and the volatilities $(\sigma_{1,1}, \sigma_{1,2}) \ldots$, $(\sigma_{n(\Delta),1}, \sigma_{n(\Delta),2})$, $(\sigma_{n(\Delta)+1,1}, \sigma_{n(\Delta)+1,2})$, equations (2.15) and (2.16) - (2.17) define a linear state-space model with state variable $x_j = \log(S_j)$ and Gaussian innovations. Hence, using a Forward-Backward algorithm to sample the true asset prices is straightforward. For Steps (4) and (5), we note that

$$\log\{|\log(S_j/S_{j-1}) - \mu(\Delta) - J_j(\Delta)|\} = \frac{1}{2}\log(\sigma_{j,1}) + \frac{1}{2}\log(\sigma_{j,2}) + \log(\epsilon_{j,1}^2)/2.$$

Following Omori et al. [2007] we approximate the error term using a mixture of Gaussian distributions,

$$\log(\epsilon_{j,1}^2)/2 \sim \sum_{l=1}^{10} p_l N\left(\frac{m_l}{2}, \frac{v_l^2}{4}\right)$$

(see Table 2.2 for the values of $\{p_l\}$, $\{m_l\}$ and $\{v_l\}$). The mixture can be rewritten

Component	p_l	m_l	v_l^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	-0.85173	0.62699
6	0.18842	-1.97278	0.98583
7	0.12047	-3.46788	1.57469
8	0.05591	-5.55246	2.54498
9	0.01575	-8.68384	4.16591
10	0.00115	-14.65000	7.33342

Table 2.2: Parameters of the mixture representation of the log Chi-squared distribution, provided in Omori et al. [2007].

by introducing auxiliary indicators $\gamma_1, \ldots, \gamma_{n(\Delta)}$ such that

$$\log(\epsilon_{j,1}^2)/2 \mid \gamma_j \sim N\left(\frac{m_{\gamma_j}}{2}, \frac{v_{\gamma_j}^2}{4}\right), \qquad \Pr(\gamma_k = l) = p_l.$$

The auxiliary indicators are sampled jointly conditional on all other parameters, and $\Pr(\gamma_k = l) = p_l$ is interpreted as the prior probability that observation k belongs to mixture element l.

Conditionally on the true prices and the indicators $\gamma_1, \ldots, \gamma_{n(\Delta)}$, we have again a linear state-space model with Gaussian innovations so that we can integrate out the volatilities in the posterior and sample all of the volatility parameters except for ρ to complete Step (3). Conditional on this sample, the volatilities can be sampled using another Forward-Backward algorithm for Step (4). Details of the algorithm are given in Appendix A.

Once the algorithm has converged and the burn-in samples have been discarded, point and interval estimates can be easily obtained using empirical estimates. For example, given a sample of the volatility path $((\sigma_{1,1}^{(b)}, \sigma_{1,2}^{(b)}), (\sigma_{2,1}^{(b)}, \sigma_{2,2}^{(b)}), \ldots, (\sigma_{n(\Delta),1}^{(b)}, \sigma_{n(\Delta),2}^{(b)}))$ for $b = 1, \ldots, B$, a sample of the in-sample (approximate) integrated variance $IV = \int_0^T \hat{\sigma}_t^2 dt$ can be obtained as

$$IV^{(b)} \approx \sum_{j=1}^{n(\Delta)} \left(\sigma_{j,1}^{(b)} \sigma_{j,2}^{(b)} \right).$$

A similar approach can be used to make out-of-sample predictions of the integrated volatility.

2.5 The effect of the mean-reverting rate $\hat{\theta}$ and the observational duration on the posterior variance of the mean log-volatility $\hat{\alpha}$

When estimating model parameters, the common intuition is that an increase in sample size leads to a decrease in posterior uncertainty. When dealing with the estimation of stochastic volatility models for high-frequency data, one may apply this thinking when the sample size is increased by obtaining more frequent price path samples for a fixed observational period, *i.e.* infill asymptotics. However, in the case where the volatility process has a finite non-zero mean-reversion timescale (as is the case for the Ornstein-Uhlenbeck processes used here), an increase in the number of intraperiod observations does not add information about the mean-level of the process. Rather, the posterior uncertainty for this model parameter can only be decreased by increasing how long we observe the process, *i.e.* increasing domain asymptotics.

To demonstrate this feature of the model, we study analytically the relationship between the mean-reverting rate $\hat{\theta}$, the time duration of observation T, and the posterior variance of mean log-volatility $\hat{\alpha}$ for a single volatility process. To proceed analytically, we consider a simplified inference problem described by the following assumptions: (1) the mean log-volatility $\hat{\alpha}$ is the only parameter to be inferred – all other parameters are known; (2) the prior distribution for the mean log-volatility $\hat{\alpha}$ is normal and is denoted by $N(\hat{a}_{\hat{\alpha}}, \hat{b}_{\hat{\alpha}}^2)$ (previously there was no parametric assumption made on the on the prior for $\hat{\alpha}$); c) the log-volatility $\log(\hat{\sigma}_t)$ is observed exactly (without error) on a uniform grid $\{0, \Delta, 2\Delta, \ldots, N\Delta\}$ in time duration $[0, T]$ where Δ is the sampling period and $N = T/\Delta$.

The exact solution of the Ornstein-Uhlenbeck process (2.2) is given in (2.7). Applying the exact solution (2.7) to the time interval $[j\Delta, (j+1)\Delta]$, we obtain

$$\log(\hat{\sigma}_{(j+1)\Delta}) = \theta(\Delta)\log(\hat{\sigma}_{j\Delta}) + (1-\theta(\Delta))\hat{\alpha} + \tau(\Delta)\varepsilon_j \;\;, \qquad 0 \le j \le N-1$$

where $\varepsilon_j \sim N(0,1)$, and $\theta(\Delta)$ and $\tau(\Delta)$ are given in (2.13). Recall the stationary distribution of the continuous-time log-volatility process in (2.8)

$$\log(\hat{\sigma}_t) \sim N\left(\hat{\alpha}, \tau(\infty)^2\right), \qquad \tau(\infty)^2 = \frac{\hat{\tau}^2}{2\hat{\theta}}.$$

Here we denote the stationary variance as $\tau(\infty)^2$ for mathematical convenience. The likelihood of $\hat{\alpha}$ given the observation $\{\log(\hat{\sigma}_0), \log(\hat{\sigma}_1), \ldots, \log(\hat{\sigma}_N)\}$ is

$$\begin{aligned} L\left(\hat{\alpha}\,|\log(\hat{\sigma}_0), \log(\hat{\sigma}_1), \ldots, \log(\hat{\sigma}_N)\right) &\propto \exp\left(\frac{-(\log(\hat{\sigma}_0)-\hat{\alpha})^2}{2\tau(\infty)^2}\right) \\ &\times \prod_{j-0}^{N-1} \exp\left(\frac{-\left(\log\left(\hat{\sigma}_{(j+1)\Delta}\right) - \theta(\Delta)\log(\hat{\sigma}_{j\Delta}) - (1-\theta(\Delta))\hat{\alpha}\right)^2}{2\tau(\Delta)^2}\right). \end{aligned} \tag{2.25}$$

Since we assume $N(\hat{a}_{\hat{\alpha}}, \hat{b}_{\hat{\alpha}}^2)$ as the prior for $\hat{\alpha}$, the posterior distribution of $\hat{\alpha}$ is normal and the reciprocal of the posterior variance of $\hat{\alpha}$ has the expression

$$\frac{1}{\mathrm{Var}\left[\hat{\alpha}\right]} = \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{1}{\tau(\infty)^2} + \sum_{j=0}^{N-1} \frac{(1-\theta(\Delta))^2}{\tau(\Delta)^2} = \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{2\hat{\theta}}{\hat{\tau}^2}\left(1 + N \cdot \tanh\left(\frac{\hat{\theta}\Delta}{2}\right)\right). \tag{2.26}$$

In the above, we have used the expressions of $\tau(\Delta)$ and $\theta(\Delta)$ given in (2.13).

Now, using the linear approximation $\tanh\left(\frac{x}{2}\right) \approx \frac{x}{2}$ and setting $T = N\Delta$, we can write (2.26) as

$$\frac{1}{\mathrm{Var}\left[\hat{\alpha}\right]} \approx \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{2\hat{\theta}}{\hat{\tau}^2}\left(1 + \frac{\hat{\theta}T}{2}\right). \tag{2.27}$$

This expression is valid for $\hat{\theta}\Delta \leq 1$, i.e. when the timescale of inertia of the log-volatility process is *greater* than the spacing between observations. The important consequence of (2.27) is that decreasing Δ does not decrease the posterior variance of $\hat{\alpha}$. In other words, an increase in the number of intraperiod observations does not add information about $\hat{\alpha}$. Rather, the posterior uncertainty for $\hat{\alpha}$ can only be decreased by increasing T (increasing how long we observe the process) or increasing $\hat{\theta}$ (on average, increasing the number of reversions to the mean). The rate of information increase for $\hat{\alpha}$ with respect to T and $\hat{\theta}$ is examined under two conditions.

When $\hat{\tau}$ is fixed, $1/\mathrm{Var}\left[\hat{\alpha}\right]$ increases linearly with the time duration T and increases quadratically with the mean-reverting rate $\hat{\theta}$. The quadratic increase of $1/\mathrm{Var}\left[\hat{\alpha}\right]$ with respect to $\hat{\theta}$ is the combined result from two contributions: i) for larger $\hat{\theta}$, the variance of $\log(\hat{\sigma}_t)$ is smaller and consequently each data point is a more accurate approximation to $\hat{\alpha}$; and ii) for larger $\hat{\theta}$, the time duration $[0, T]$ covers more rounds of $\log(\hat{\sigma}_t)$ fluctuating away from $\hat{\alpha}$ and relaxing back toward $\hat{\alpha}$.

When $\frac{\hat{\tau}^2}{2\hat{\theta}}$ (the stationary variance of log-volatility) is fixed, $1/\mathrm{Var}\left[\hat{\alpha}\right]$ increases linearly with $\hat{\theta}T$. In this case, if the prior is wider than the stationary distribution ($\hat{b}_{\hat{\alpha}}^2 \geq \frac{\hat{\tau}^2}{2\hat{\theta}}$) and the time duration is much larger than the time scale of inertia ($T \gg 1/\hat{\theta}$), then the posterior variance of $\hat{\alpha}$ is inversely proportional to the time duration:

$$\mathrm{Var}\left[\hat{\alpha}\right] \approx \frac{\hat{\tau}^2}{\hat{\theta}} \cdot \frac{1}{\hat{\theta}T}. \tag{2.28}$$

When $\hat{\theta}\Delta \gg 1$ (i.e. when the timescale of inertia of the log-volatility process is much *smaller* than the spacing between observations), the linear approximation $\tanh(\hat{\theta}\Delta/2) \approx \hat{\theta}\Delta/2$ is invalid. Instead, we have the approximation $\tanh(\hat{\theta}\Delta/2) \approx 1$, which leads to

$$\frac{1}{\mathrm{Var}\left[\hat{\alpha}\right]} \approx \frac{1}{\hat{b}_{\hat{\alpha}}^2} + \frac{2\hat{\theta}}{\hat{\tau}^2}(1+N), \qquad \text{for } \hat{\theta}\Delta \gg 1$$

Under this regime, $1/\mathrm{Var}\left[\hat{\alpha}\right]$ is approximately proportional to the number of observations, N, provided that $\hat{\theta}\Delta \gg 1$ is preserved as N is increased. This occurs when the spacing between observations, Δ, is fixed and the increase in N comes from extending the observational duration T. When T is fixed, as N increases Δ decreases, which eventually will carry the system from the regime of $\hat{\theta}\Delta \gg 1$ to that of $\hat{\theta}\Delta \leq 1$. The behavior of the posterior variance of $\hat{\alpha}$ for $\hat{\theta}\Delta \leq 1$ with N increasing through either lowering Δ or increasing T is illustrated in Section 2.6.4 below.

The results above, which indicate that $\hat{\alpha}$ cannot be estimated consistently under in-fill asymptotics, apply to any stochastic volatility model based on the Ornstein-Uhlenbeck process. Note, however, that they do not contradict standard asymptotic results from the realized volatility literature, which focus on the integrated variance during a finite period of time and not on the long-term median volatility of the process.

2.6 Illustrations

2.6.1 Effect of microstructure noise and sampling frequency on estimates: simulation studies

In this section we examine the effects of microstructure noise and sampling frequency on our inference of the model parameters. We first consider a one-day (6.5 hours) simulated dataset in which the true log-prices were generated according to (2.16) - (2.17) with $\Delta = 10^{-3}$ seconds, but where the microstructure noise was incorporated by adding to each point (exponentiated to transform from log-price to price level) a uniformly distributed random variable between -0.05 and 0.05, simulating a \$0.1 bid-ask spread. The price was rounded to the nearest 100th, then transformed back to the log scale.

The true parameters used in the simulations were set to be reasonably close to typical values on the S&P500 market. The instantaneous return per millisecond $\hat{\mu}$ was set to $\hat{\mu} = 1.7 \cdot 10^{-12}$/millisecond, corresponding to an annual return of 1%, based on 251 trading days per year, 6.5 trading hours per trading day, excluding jumps between trading sessions. Assuming a characteristic timescale of inertia of the slow continuous log-volatility process to be 3.5 hours (measured in milliseconds) and that of the fast log-volatility process to be 10 minutes, $\hat{\theta}_1$ and $\hat{\theta}_2$ were set to $\hat{\theta}_1 = 7.94 \cdot 10^{-8}$/milliseconds and $\hat{\theta}_2 = 1.57 \cdot 10^{-6}$/milliseconds. The remaining parameters $\hat{\tau}_i^2$ and $\hat{\alpha}$ governing the behavior of the log-volatility processes were set using the publicly traded VIX index. The VIX is the square root of the risk-neutral market expectation of the S&P 500 variance over the next 30 days on an annualized scale, such that

$$\log \hat{\sigma}_t \approx \log\left(\log\left(1 + \mathrm{VIX}_t/100\right)\right) - \frac{1}{2}\log(T_{\mathrm{year}})$$

where $T_{\text{year}} = 1$ year $= 251 \times 6.5 \times 3.6 \cdot 10^6$ ms. We can transform between VIX_t and $\log(\hat{\sigma}_t)$, obtaining approximations of the historical log-volatility path. With $\hat{\theta}_i$ set, the parameters $\hat{\alpha}$ and $\hat{\tau}_i^2$ determine the stationary distribution of the log-volatility process, $\log(\hat{\sigma}_{t,i}) \sim N\left(\hat{\alpha}, \frac{\hat{\tau}_i^2}{2\hat{\theta}_i}\right)$. Thus we set $\hat{\alpha}$ to be the mean of this VIX-derived log-volatility path, and $\hat{\tau}_i^2$ the variance thereof, divided by $2\hat{\theta}_i$. A summary of the true model parameters used in this simulation is presented in Table 2.3.

We fit our model to the simulated data using priors whose means equal the true simulation values and whose standard deviations put the majority of the prior mass within one order of magnitude of the mean (recall that the prior specification procedure described in Section 2.3 only requires the first two moments of each continuous-time parameter). For example, since $\hat{\mu}_T = 1.7 \cdot 10^{-12}$, we let $\text{E}[\hat{\mu}] = \hat{a}_{\hat{\mu}} = 1.7 \cdot 10^{-12}$ and $\text{Var}[\hat{\mu}] = \hat{b}_{\hat{\mu}}^2 = (1 \cdot 10^{-11})^2$. In this case, the prior region covered by three standard deviations to the left and right of the prior mean approximately corresponds to a range from -15% to a 20% annual return. In general, specifying the prior standard deviation in this manner usually leads to a relatively wide but reasonable prior coverage of parameter model values. A summary of the prior means and standard deviations used to fit the model is provided in Table 2.4.

We fit three slightly different versions of our model to the simulated dataset. In the first version, the microstructure noise parameter was set to $\xi^2 = 0$, so our model does not take into account microstructure noise and reduces to the standard SV models used in the literature. In the second version, ξ^2 was fixed to $2.5 \cdot 10^{-7}$, a level of microstructure noise roughly consistent with the true level of microstructure noise added in the data. Finally, the third version corresponds to our full model where ξ^2 is estimated from the data by assigning it a Gamma prior with mean $2.5 \cdot 10^{-7}$ and standard deviation $1 \cdot 10^{-6}$ (see Table 2.4). The dataset

Parameter	Value	Interpretation
$\hat{\mu}$	$1.7 \cdot 10^{-12}/\text{ms}$	Annual asset return of 1%, based on 251 trading days per year, 6.5 trading hours per trading day, excluding jumps between trading sessions.
$\hat{\theta}_1$	$7.94 \cdot 10^{-8}/\text{ms}$	Timescale of inertia for the slow log-volatility process equal to 3.5 hours.
$\hat{\theta}_2$	$1.67 \cdot 10^{-6}/\text{ms}$	Timescale of inertia for the slow log-volatility process equal to 10 minutes.
$\hat{\alpha}$	$-13 - \log(\sqrt{\text{ms}})$	Average of daily closing VIX values from 1/2/1990 to 4/10/2015 (18.9% a year), transformed to the $\log(\hat{\sigma}_t)$ scale with time measured in milliseconds.
$\hat{\tau}_1^2$	$1.86 \cdot 10^{-8}/\text{ms}$	$2\hat{\theta}_1$ times the variance of daily closing VIX values from 1/2/1990 to 4/10/2015, transformed to the $\log(\hat{\sigma}_t)$ scale with time measured in milliseconds. In terms of the annualized volatility for the price, the corresponding distribution of annualized volatility has (1st, 10th, 50th, 90th, 99th) percentiles given by $(8.16\%, 11.8\%, 18.9\%, 30.7\%, 46.6\%)$
$\hat{\tau}_2^2$	$3.9 \cdot 10^{-7}/\text{ms}$	$2\hat{\theta}_2$ times the variance of daily closing VIX values from 1/2/1990 to 4/10/2015, transformed to the $\log(\hat{\sigma}_t)$ scale with time measured in milliseconds. In terms of the annualized volatility for the price, the corresponding distribution of annualized volatility has (1st, 10th, 50th, 90th, 99th) percentiles given by $(8.16\%, 11.8\%, 18.9\%, 30.7\%, 46.6\%)$
ξ_T^2	$2.5 \cdot 10^{-7}$	Bid-ask spread of \$0.1 for an average price of \$100.
ρ_T	0	Innovations in the price and log-volatility process are independent, no leverage effect.

Table 2.3: Summary for model parameters in simulation data, along with associated market interpretation of these parameters.

was analyzed assuming sampling periods of 300, 30, 15, and 5 seconds (note that, because the size of the microstructure noise is assumed to be the same at every sampling scale, we use the same prior for ξ^2 for all sampling periods).

Parameter	Prior mean	Prior standard deviation
$\hat{\mu}$	$1.7 \cdot 10^{-12}$	$1 \cdot 10^{-11}$
$\hat{\theta}_1$	$7.94 \cdot 10^{-8}$	$1 \cdot 10^{-7}$
$\hat{\theta}_2$	$1.67 \cdot 10^{-6}$	$1 \cdot 10^{-5}$
$\hat{\tau}_1^2$	$1.86 \cdot 10^{-8}$	$1 \cdot 10^{-7}$
$\hat{\tau}_2^2$	$3.9 \cdot 10^{-7}$	$1 \cdot 10^{-6}$
$\hat{\alpha}$	-13	10
ξ^2	$2.5 \cdot 10^{-7}$	$1 \cdot 10^{-6}$
ρ	0	1

Table 2.4: Parameters of the prior distributions used for inference with simulated data.

The plots for the posterior mean and 95% probability bands for the estimated log-volatility paths are given in Figure 2.1 - 2.2, along with the true signal from the simulated data. As the sampling period decreases, we see that the first version of the model (which ignores microstructure noise by fixing $\xi^2 = 0$) fails to capture the latent signal. On the other hand, the other two versions of our models produce much better posterior estimates for the latent signal. In particular, we see that, as expected, the naive choice $\xi^2 = 0$ model overestimates the volatility signal for higher sampling frequencies where the time interval Δ diminishes enough so that the microstructure noise dominates for the volatility signal. We can also see that, although the second and third versions of the model tend to smooth out the true volatility path, the reconstruction generated by the model that estimates ξ^2 from the data is somewhat more accurate.

In addition to estimating the volatility path, we also investigate the ability of the model to infer model parameters. In particular, Figure 2.4 shows the posterior density estimates for the continuous-time parameters $\hat{\alpha}$, $\hat{\tau}^2$, $\hat{\theta}$ and $\hat{\mu}$ (which are comparable across scales), as well as the posterior distribution for ξ^2 (in the case of the third version of the model, which is the only one in which it is estimated from the data). Note that, when the model is estimated with $\xi^2 = 0$ fixed,

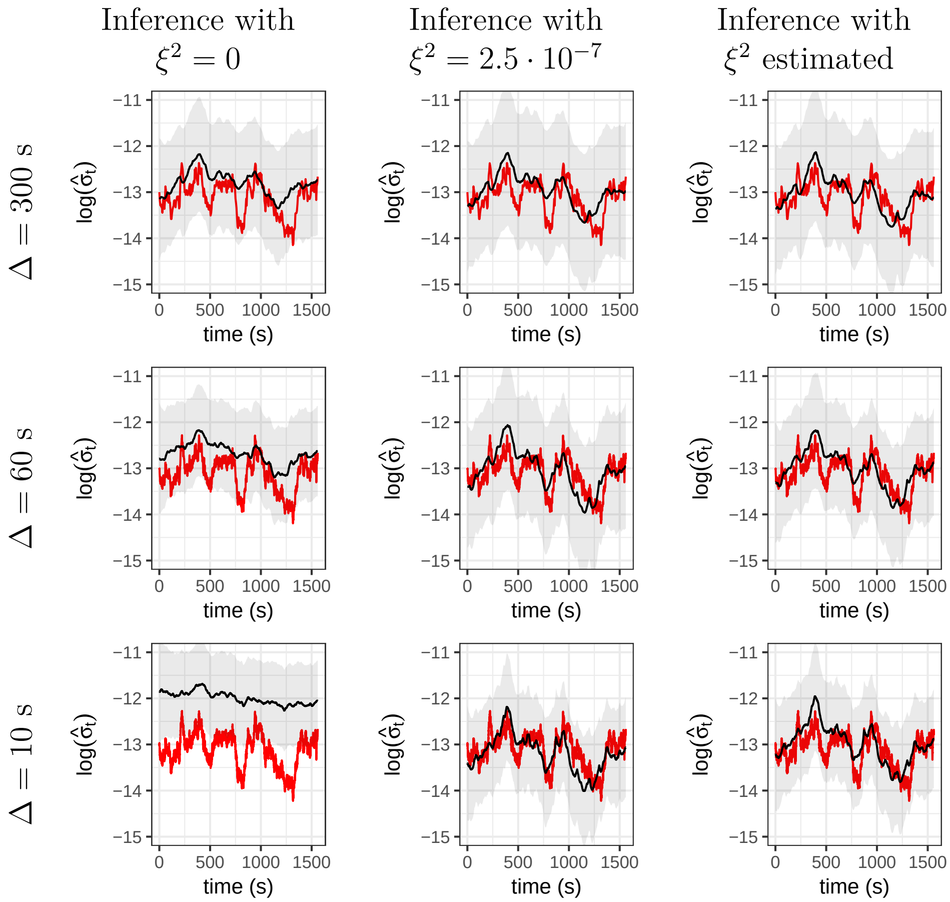

Figure 2.1: Slow log-volatility paths for simulated data. All three inference approaches are applied to the same data set that contains microstructure noise. The microstructure noise added in the simulated data is approximately at the level of $\xi^2 = 2.5 \cdot 10^{-7}$. Red denotes the true paths, while the gray region denotes the posterior 95% probability for the log-volatility value. Black is the posterior mean log-volatility signal. We see that when microstructure is ignored ($\xi^2 = 0$), we fail to recover the true signal.

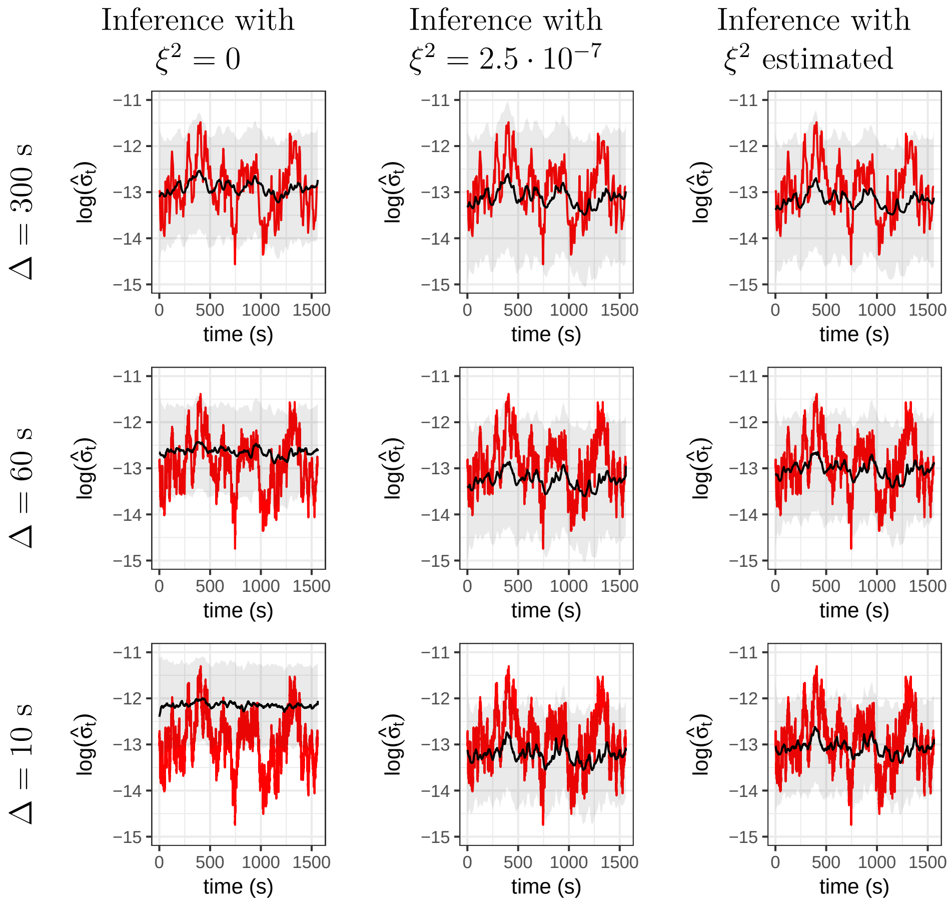

Figure 2.2: Fast log-volatility paths for simulated data. All three inference approaches are applied to the same data set that contains microstructure noise. The microstructure noise added in the simulated data is approximately at the level of $\xi^2 = 2.5 \cdot 10^{-7}$. Red denotes the true paths, while the gray region denotes the posterior 95% probability for the log-volatility value. Black is the posterior mean log-volatility signal. We see that when microstructure is ignored ($\xi^2 = 0$), we fail to recover the true signal.

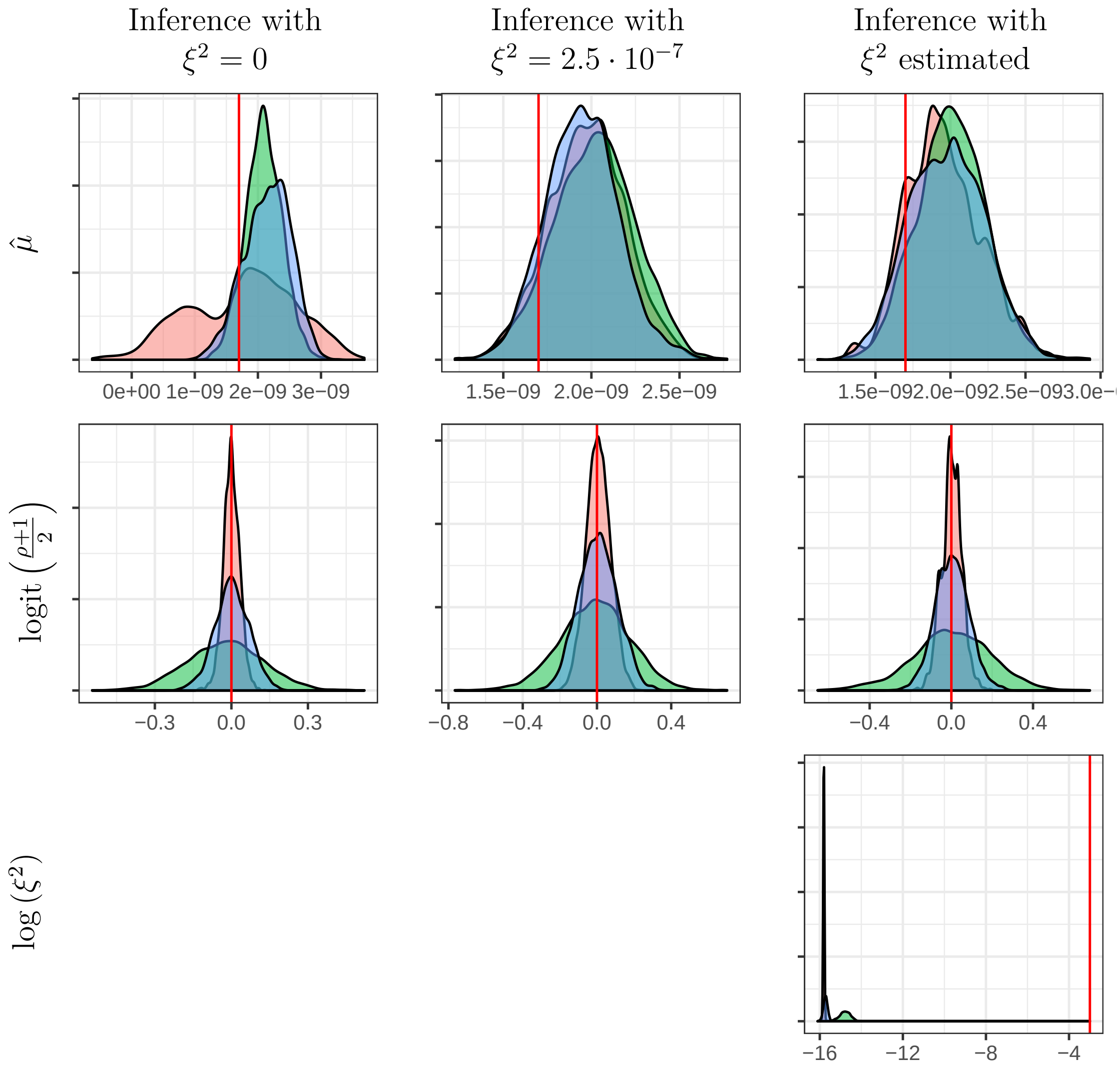

Figure 2.3: Posterior density approximations of the observational model parameters for simulated data. The red vertical line represents the true parameter value. The sampling periods used are: 5 minutes (red), 60 seconds (blue), and 10 seconds (red).

the posterior densities for $\hat{\alpha}$, the mean level of log volatility, show a reduction in variance with increasing sampling frequency: posterior draws become more centered around a wrong, overestimated value for mean log-volatility level. These results are consistent with those obtained for the volatility path and show that the model fails to capture the constant information content in the data regarding

$\hat{\alpha}$. We also note that learning $\hat{\tau}^2$ and $\hat{\theta}$ is difficult whether we do or do not include microstructure noise. However, due to the constant information in the data with respect to $\hat{\alpha}$, the posterior uncertainty for $\hat{\theta}$ and $\hat{\tau}^2$ seems to remain constant even with increasing sampling frequency when ξ is not fixed.

2.6.2 Estimating Integrated Variance

As described in Section 2.4, the posterior draws for σ_j^2 allow us to approximate the posterior distributions for the integrated variance of the latent volatility process. In this Section we extend the previous simulation study to compare the 95% intervals generated by the three versions of our model with those generated from a realized variance estimate. The literature on realized volatility estimators for high-frequency data is vast (for a review, refer to Pigorsch et al. [2012]), but the construction of confidence intervals for the realized volatility estimators can be challenging. Here we compare the coverage properties of our model-based credible intervals against bootstrap-based confidence intervals of the the kernel-based realized variance estimator introduced by Zhou [1996] and Hansen and Lunde [2006]. The idea behind the bootstrapping method is to periodically extend the available data set and randomly re-select a new data set to construct a bootstrap sample (see Hwang et al. [2013] for a full description of the procedure).

The results of the comparison are shown in Table 2.5. The table is constructed using 300 simulated data sets, each corresponding to a single trading day. We compare the percentage of times the 95% confidence/credible intervals for the integrated variance (IV) estimator covers the true integrated variance value. A well-calibrated interval will produce a 95% coverage on average, and we see that the estimator based on our approach where ξ^2 is fully estimated preforms very well, both when compared to $\xi^2 = 0$, $\xi^2 = 2.5 \cdot 10^{-7}$, and when compared to the kernel-based estimator.

	Sampling period				
	5 min	60 sec	30 sec	15 sec	5 sec
Inference with $\xi^2 = 0$	93	72	28	3	0
Inference with $\xi^2 = 2.5 \cdot 10^{-7}$	95	79	57	23	0
Inference with ξ^2 estimated	95	91	92	96	97
Inference with kernel-based estimator	53	51	48	59	76

Table 2.5: Coverage table comparing the percentage of covered integrated variance levels by 95% confidence/credible intervals of our estimator and a kernel-based estimator.

2.6.3 Effect of microstructure noise and sampling frequency on estimates: market data

We perform an analysis for real market data, which consists of a single day of midpoint spot prices of Apple Inc. (NASDAQ:AAPL) on March 6^{th}, 2014, printed on the millisecond from the NYSE TAQ data set. Our a priori distribution for the bid-ask spread driving microstructure noise is centered on \$0.1 as with the simulation data. All other priors are the same as in Section 2.6. The estimated volatility paths are shown in Figure 2.5. In the case where $\xi^2 = 0$, the model estimates the volatility signal to be, on average, higher than in the cases where $\xi^2 > 0$, which is also seen in the posterior means estimates of $\hat{\alpha}$ in Figure 2.6. This is consistent with the simulation-study results, where the $\xi^2 = 0$ model attributes microstructure noise to the log-volatility signal.

Figure 2.6 shows the posterior distributions for the model parameters. Note that the posterior for ξ^2 is centered around $8 \cdot 10^{-9}$ – two orders of magnitude smaller than the prior center of $2.7 \cdot 10^{-7}$. This value of the posterior mean is roughly equivalent to a bid-ask spread of \$0.01, which is reasonable for a highlytraded stock like AAPL. Furthermore, note that fixing $\xi^2 = 2.5 \cdot 10^{-7}$ compared to treating ξ^2 as an unknown parameter, leads to volatility paths that are smoother and have greater coherence in posterior estimates of the model parameters across sampling periods. This is especially true for $\hat{\theta}$, since the closer the log-volatility process is to being discontinuous, the shorter its timescale must be. We thus see the obvious trade-off when specifying ξ^2: if ξ^2 is too large, we run the risk of over-smoothing; if ξ^2 is too small, we confound the effects of noise with those of the volatility process.

2.6.4 Effect of timescale of inertia on estimates of $\hat{\alpha}$

To illustrate the discussion in Section 2.5 on the effect of the log-volatility timescale on our method's ability to learn $\hat{\alpha}$, we examine the posterior uncertainty for $\hat{\alpha}$ under two scenarios: (1) increasing sample size N by decreasing the sampling period Δ, and (2) increasing N by increasing the observational period T while keeping Δ constant. The same simulated dataset is used in both (1) and (2), with $1/\hat{\theta} = 15$ min, such that $\hat{\theta}\Delta \leq 1$ and the approximation for the posterior variance of $\hat{\alpha}$ in (2.27) is applicable.

Under scenario (1), we consider the entire data set and estimate $\hat{\alpha}$ with $\Delta =$ 1 min, 30 sec, 15 sec, and 5 sec. The posterior distributions for $\hat{\alpha}$ are shown in the left panel of Figure 2.7. We see that the posterior uncertainty for $\hat{\alpha}$ remains the same with increasing number of intraperiod samples, as suggested by the analysis in Section 2.5. Under scenario (2), we fix $\Delta = 1$ min and increase sample sizes by increasing the observational period T, using the first 1/6 (65 min) of the data, the first 2/6 (130 min) of the data, and so on through the entirety of the data (390 min). The right panel of Figure 2.7 shows the posterior densities for $\hat{\alpha}$ under this regime. Confirming the discussion in Section 2.5, we tend to see a decreasing posterior variance for $\hat{\alpha}$ with increasing observational duration, but not when the sampling frequency increases. The important takeaway point is that a dataset covering a finite observational period contains a finite amount of information, no matter now finely the observational period is sampled.

2.7 Conclusion

In this paper we have outlined a discrete-time stochastic volatility model for high-frequency data. The model and the algorithm used to estimate it are designed to be coherent across all sampling frequencies. To this end, we elicit priors on the parameters of the continuous-time version of our model and transform them to the discrete-time scale. Both simulation and real data results show that adding the microstructure term in the standard stochastic volatility formulation allows one to fit such models to high-frequency data and extract the log-volatility signal from noisy observations. However, having a good prior estimate of the noise level is an important specification, since attributing some fluctuations in the observed log volatility to microstructure noise has a smoothing effect on the reconstructed log volatility paths. Finally, simulation studies show that the integrated variance estimator derived from our model is well-calibrated and outperforms current kernel-based realized volatility estimators.

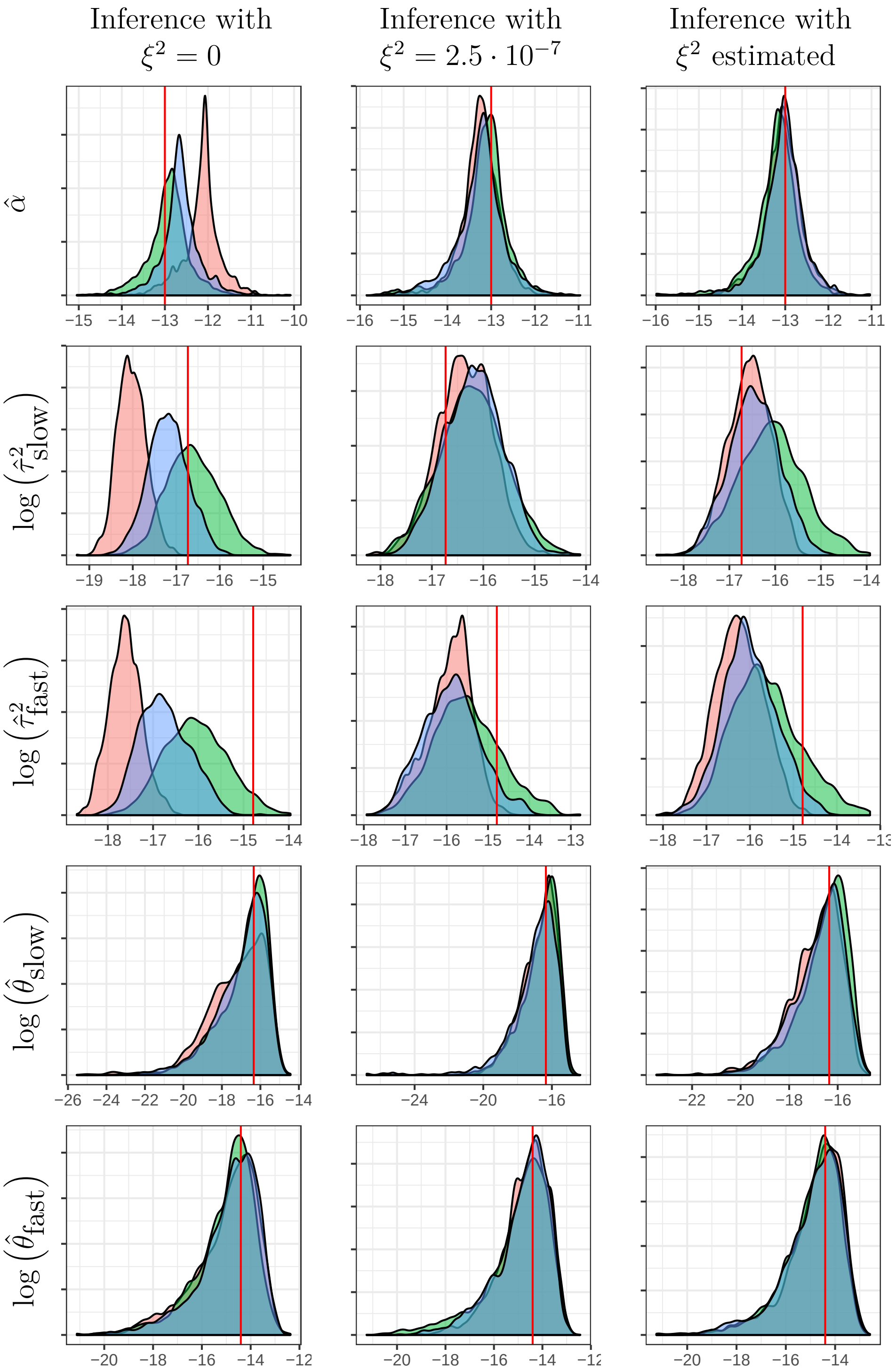

Figure 2.4: Posterior density approximations of volatility model parameters for simulated data. The red vertical line represents the true parameter value. The sampling periods used are: 5 minutes (green), 60 (blue) seconds, and 10 seconds (red).

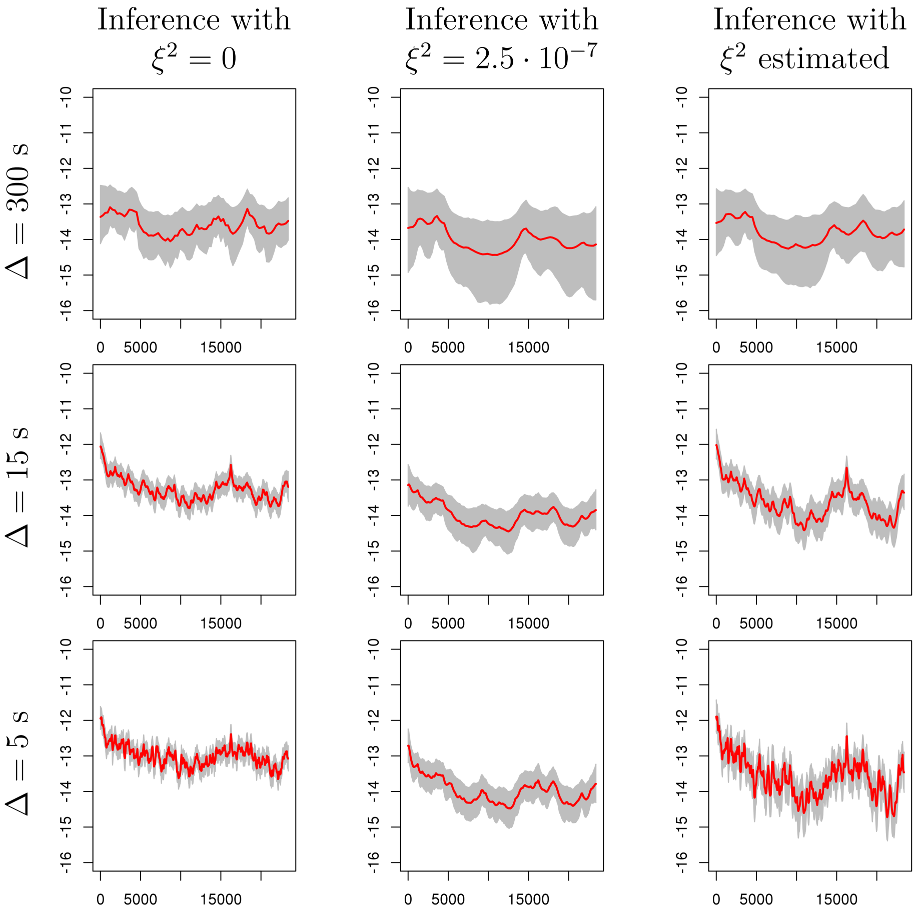

Figure 2.5: Log-volatility paths for the AAPL 03/06/2014 data. Red denotes the posterior mean of the paths, while the gray region denotes the posterior 95% probability for the log-volatility value.

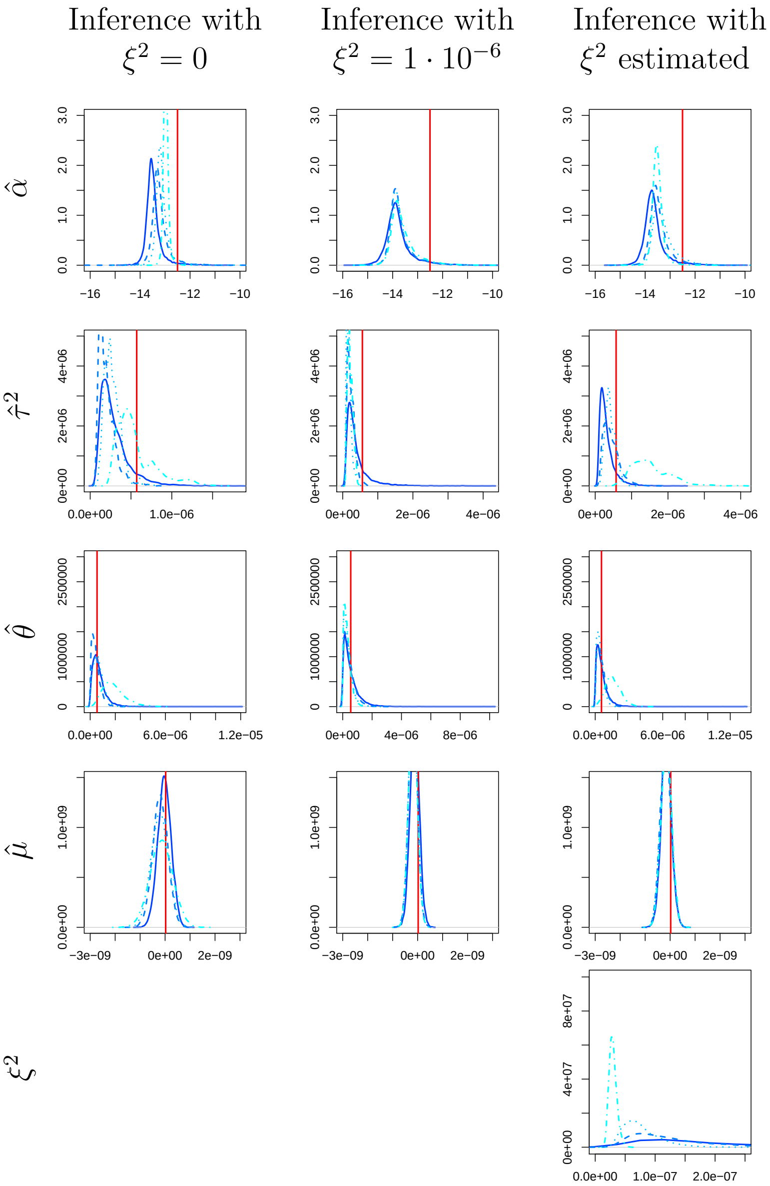

Figure 2.6: Posterior density approximations of model parameters for the AAPL 03/06/2014 data. The sampling periods used are: 5 minutes (——) , 30 seconds (– –) , 15 seconds (· · · ·) , 5 seconds (· – ·) , and 1 second (— —) . Vertical red lines represent prior mean for the given parameters.

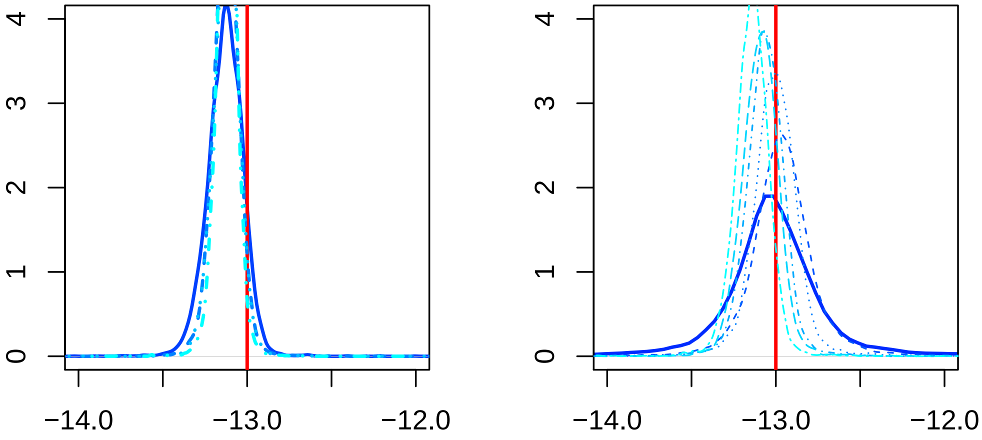

Figure 2.7: Posterior densities of $\hat{\alpha}$. Left panel: sample size N is increased by decreasing Δ. Right panel: sample size N is increased by increasing the observational duration T. The same data set was used for both cases, where $1/\hat{\theta} = 15\,\text{min}$ so that $\hat{\theta}\Delta \leq 1$ for all posterior samples. The red vertical line signifies the true $\hat{\alpha}$ value used in the data-generation process. For a fixed T and decreasing Δ (left) [1 minute (——) , 30 seconds (– –) , 15 seconds (· · · ·) , 5 seconds (· – ·)] the posterior variance of $\hat{\alpha}$ stays approximately the same. For a fixed Δ and increasing T [65 minutes (——) , 130 minutes (– –) , 195 minutes (· · · ·) , 260 minutes (· – ·) , 325 minutes (– –) , and 390 (– –) minutes], the posterior variance of $\hat{\alpha}$ tends to decrease.

Chapter 3

Volatility Estimation For Bivariate Assets With OCHL Data: A Galerkin Solution

3.1 Introduction

In Chapter 2, we showed how a filtering approach can be effectively applied to very high-frequency univariate data for the more accurate estimation of asset volatility. Information present in high-frequency returns can also be captured over longer observational periods by including the highest and lowest intraperiod prices. This information can substantially improve estimates (see for example Horst et al. [2012] for the univariate treatment). In this chapter we develop a general likelihood-based approach for incorporating open, close, high, and low data in the bivariate setting.

We consider a two-dimensional correlated Brownian motion:

$$X(t) = x_0 + \mu_x t + \sigma_x W_x(t) \tag{3.1}$$

$$Y(t) = y_0 + \mu_y t + \sigma_y W_y(t) \tag{3.2}$$

where W_i are standard Brownian motions with $\mathrm{Cov}(W_1(t), W_2(t)) = \rho t$. Our interest is in finding the 6-dimensional joint probability density function for the pair $(X(t), Y(t))$ and the random variables $M_X(t) = \max_{0\leq s\leq t} X(s)$, $m_X(t) = \min_{0\leq s\leq t} X(s)$, $M_Y(t) = \max_{0\leq s\leq t} Y(s)$, $m_Y(t) = \min_{0\leq s\leq t} Y(s)$. We introduce

$$\begin{aligned} q(x, y, t) = \lim_{dy\to 0} \lim_{dx\to 0} \frac{1}{dxdy} \mathrm{Pr}\,(x \leq X(t) < x + dx, y \leq Y(t) < y + dy, \\ M_x(t) \leq b_x, M_y(t) \leq b_y, m_x(t) \geq a_x, m_y(t) \geq a_y | X(t) = x_0, Y(0) = y_0, \theta) \end{aligned} \tag{3.3}$$

with $\theta := (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. The density satisfies the Fokker-Planck equation [Oksendal, 2013]:

$$\begin{aligned} \frac{\partial}{\partial t'} q(x, y, t') = -\mu_x \frac{\partial}{\partial x} q(x, y, t') - \mu_y \frac{\partial}{\partial y} q(x, y, t') + \\ \frac{1}{2}\sigma_x^2 \frac{\partial^2}{\partial x^2} q(x, y, t') + \rho\sigma_x\sigma_y \frac{\partial^2}{\partial x \partial y} q(x, y, t') + \frac{1}{2}\sigma_y^2 \frac{\partial^2}{\partial y^2} q(x, y, t'). \end{aligned} \tag{3.4}$$

Given that we consider the subset of $(X(t), Y(t))$ that satisfies the constraints imposed by the boundary data, we impose the absorbing conditions on the Fokker-Planck equation

$$q(a_x, y, t') = q(b_x, y, t') = q(x, a_y, t') = q(x, b_y, t') = 0, \qquad 0 < t' \leq t. \tag{3.5}$$

Differentiating $q(x, y, t)$ with respect to the boundaries produces the 6-dimensional transition density of

$$(X(0), Y(0), M_X(t), m_X(t), M_Y(t), m_Y(t)) = (x, y, b_x, a_x, b_y, a_y)$$

given the starting point of $(X(0), Y(0)) = (x_0, y_0)$. We denote this 6-dimensional transition density by $f(x, y, a_x, b_x, a_y, b_y)$ and note that it is the fourth-order partial derivative of the unconstrained process q with respect to the four boundaries:

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q(x, y, t) = f(x, y, a_x, b_x, a_y, b_y). \tag{3.6}$$

We should note that $f(x, y, a_x, b_x, a_y, b_y)$ exists. Consider the relation

$$\begin{aligned} &f(x, y, a_x, b_x, a_y, b_y) = \\ &\qquad \Pr\left(X(t) \in dx, Y(t) \in dy, m_x \in da_x, M_x \in db_x, m_y \in a_y, M_y \in db_y \,|\theta\,\right) = \\ &\qquad\qquad \mathbb{P}_W\left(A(x, y, a_x, b_x, a_y, b_y)\right), \end{aligned}$$

where

$$\begin{aligned} A(x, y, a_x, b_x, a_y, b_y) = \Bigg\{ \omega \in W \,\Bigg|\, & X_\omega(t) = x, Y_\omega(t) = y, \inf_{t' \in [0,1]} X_\omega(t') = a_x, \\ & \sup_{t' \in [0,1]} X_\omega(t') = b_x, \inf_{t' \in [0,1]} Y_\omega(t') = a_y, \sup_{t' \in [0,1]} Y_\omega(t') = b_y \Bigg\}. \end{aligned}$$

We use the shorthand $X(t) \in dx$ for $X(t) \in (x, x + dx)$, and $\mathbb{P}_W$ is the Wiener measure on the sample space W of all realizations (paths) $(X_\omega(t), Y_\omega(t))$ from the stochastic process (3.1) - (3.2). The process is defined in the usual way using Kolmogorov's extension of measure over cylinder sets on $t \to \mathbb{R}^2$ (see Freidlin

[1985], Section 1.2). Sets of the form $A(x, y, a_x, b_x, a_y, b_y)$ can be defined as a countable intersection/union of cylinder sets on $t \to \mathbb{R}^2$, hence they are measurable under $\mathbb{P}_W$; therefore $f(x, y, a_x, b_x, a_y, b_y)$ exists.

Marginals of (3.3) involving less than all four boundaries have 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], and to pricing financial derivative instruments whose payoff depends on some (but not all) of the observed boundaries [He et al., 1998]. One-dimensional analog of this problem was used in Horst et al. [2012] to derive a full likelihood-based (Bayesian) approach to estimate volatility in *univariate* financial time series 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 chapter we will provide an efficient numerical method necessary for carrying out inferential procedures with correlated financial time series in the bivariate setting.

Closed-form solutions to (3.4) - (3.5) are only 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. For $\rho \neq 0$, the solution when $a_x = -\infty$ and $b_x = \infty$, can be obtained using the method of images to enforce the remaining boundaries. Alternatively, if $a_x, a_y = -\infty$ or $b_x, b_y = \infty$, eigenfunctions of the Fokker-Plank equation can be found in radial coordinates. Both of these techniques are used and detailed by He et al. [1998]. A further complication is that we are not directly interested in evaluating the solution of the differential equation, but we are interested instead in evaluating its fourth order mixed partial derivative (recall equation (3.6)).

For non-zero correlation, it is still possible to approach the general problem by proposing a biorthogonal expansion in time and space [Risken, 1989, section 6.2], where the eigenfunctions for the differential operator are approximated by the eigenvectors of a linear system obtained using a truncated expansion based on eigenfunctions for the case $\rho = 0$, a set of separable basis functions, each of which is a product of two sine functions (one in each dimension) 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 moderately large values of $|\rho|$, a large number of basis elements is typically needed for an accurate approximation. The denseness and size of the resultant system makes the expansion a slow solution. This becomes unfeasible in cases where we need to solve the equation many times such as in the case of doing statistical likelihood-based inference. 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. The trigonometric series expansion has exponential accuracy in the case $\rho = 0$, while it loses accuracy catastrophically when $\rho \neq 0$. We hypothesize that the inefficiencies in the trigonometric expansion arise from using a separable representation for a differential operator that is intrinsically correlated in the two dimensions. For the finite difference approach, the problem is twofold as the method introduces numerical diffusion in the initial condition and it also uses a separable approximation for the differential operator.

In this chapter we propose a robust and efficient algorithm to numerically approximate the solution to (3.4) - (3.5) and (3.6) in two distinct parameter regimes. We use a Galerkin discretization that directly takes into account the correlation parameter present in the differential operator in order to efficiently represent an analytic small-time solution and propagate it forward in time. We apply our computational method to estimate diffusion parameters with a maximum likelihood from an independent, identically distributed sample. Section 3.2 outlines some methods we considered for this problem. Section 3.4 describes the Galerkin method we implemented. Section 3.5.3 includes our numerical experiments.

3.2 Preliminaries

Before considering solutions to the full Fokker-Planck equation (3.4) - (3.5), we simplify the PDE by proposing a series of transformations to standardize the problem. Firstly, transformation

$$T_{(0)} : q(x, y, t) \to p(x, y, t) \tag{3.7}$$

decomposes solution q into an exponential factor and an unknown function p, eliminates the advection terms in the differential equation for p while preserving the initial and boundary conditions for p. $T_{(0)}$ is defined as

$$T_{(0)} : p(x, y, t) = \exp(\alpha x + \beta y + \gamma t) q\left(x,\ y,\ t\right), \tag{3.8}$$

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 functions satisfies the *pure diffusion* equation:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sigma_x^2\frac{\partial^2 p}{\partial x^2} + \rho\sigma_x\sigma_y\frac{\partial^2 p}{\partial x\partial y} + \frac{1}{2}\sigma_y^2\frac{\partial^2 p}{\partial y^2}, \qquad (x,y)\in\Omega \tag{3.9}$$

$$\Omega := (a_x, b_x)\times(a_y, b_y)$$

subject to the constraints

$$p(x,y,t) = 0, \text{ for } (x,y)\in\partial\Omega,$$

$$p(x,y,0) = \delta\left(x-x_0\right)\delta\left(y-y_0\right).$$

Secondly, the problem is then *standardized* with the scaling transformation

$$T_{(1)} : \begin{cases} p_{(1)}(x^{(1)}, y^{(1)}, t) & = & L_x L_y p(x,y,t), \\ p_{(1)}(x^{(1)}, y^{(1)}, 0) & = & \delta(x^{(1)} - x_0^{(1)})\delta(y^{(1)} - y_0^{(1)}), \\ (x^{(1)}, y^{(1)}) & = & \left(\frac{x-a_x}{L_x}, \frac{y-a_y}{L_y}\right), \\ (L_x, L_y) & = & (b_y - a_y, b_x - a_x) \end{cases} \tag{3.10}$$

which re-defines the differential equation and computational domain observed by $p_{(1)}(x^{(1)}, y^{(1)}, t)$

$$\frac{\partial}{\partial t}p_{(1)}(x^{(1)}, y^{(1)}, t) = \mathcal{L}^{(1)}p_{(1)}(x^{(1)}, y^{(1)}, t), \quad (x^{(1)}, y^{(1)})\in\Omega^{(1)}, \tag{3.11}$$

$$\Omega^{(1)} := (0,1)\times(0,1).$$

The differential operator $\mathcal{L}^{(1)}$ takes the form

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

with $\tau_x = \frac{\sigma_x}{L_x}$, $\tau_y = \frac{\sigma_y}{L_y}$. Two remarks are in order here:

1. The transformation $T_{(1)}$ preserves ρ from the original problem.
2. We consider L_x and L_y as *fixed* in differentiating $p(x, y, t)$, corresponding to the unperturbed (a_x, b_x, a_y, b_y). This amounts to perturbing the boundaries of $\Omega^{(1)}$ when differentiating $p_{(1)}(x^{(1)}, y^{(1)}, t)$ with respect to (a_x, b_x, a_y, b_y), while keeping the diffusion parameters τ_x, τ_y constant. In contrast, if we treat L_x, L_y as functions of the boundary parameters, then upon differentiation τ_x, τ_y will be perturbed while preserving $\Omega^{(1)}$.

Thirdly, we *normalize* the diffusion equation (3.11) with the transformation

$$T_{(2)} : p_{(2)}(\tilde{x}, \tilde{y}, \tilde{t}) = \begin{cases} p_{(1)}(x^{(1)}, y^{(1)}, \tilde{t}/\tau_x^2), & \text{if } \tau_x^2 \geq \tau_y^2 \\ p_{(1)}(y^{(1)}, x^{(1)}, \tilde{t}/\tau_y^2), & \text{otherwise,} \end{cases} , \tag{3.12}$$

defined with the coordinate transformation

$$\tilde{x} = x^{(1)} \cdot \mathbb{1}_{\left(\max\left\{\tau_x^2, \tau_y^2\right\} = \tau_x^2\right)} + y^{(1)} \cdot \mathbb{1}_{\left(\max\left\{\tau_x^2, \tau_y^2\right\} = \tau_y^2\right)}, \tag{3.13}$$

$$\tilde{y} = y^{(1)} \cdot \mathbb{1}_{\left(\max\left\{\tau_x^2, \tau_y^2\right\} = \tau_x^2\right)} + x^{(1)} \cdot \mathbb{1}_{\left(\max\left\{\tau_x^2, \tau_y^2\right\} = \tau_y^2\right)}, \tag{3.14}$$

$$\tilde{t} = t \cdot \max\left\{\tau_x^2, \tau_y^2\right\}. \tag{3.15}$$

The differential problem under $T_{(2)}$ now takes on the *normalized form* (from hereon referred to as the *normalized problem*):

$$\frac{\partial}{\partial \tilde{t}} p_{(2)}(\tilde{x}, \tilde{y}, \tilde{t}) = \tilde{\mathcal{L}} p_{(2)}(\tilde{x}, \tilde{y}, \tilde{y}), \qquad (\tilde{x}, \tilde{y}) \in \tilde{\Omega} \tag{3.16}$$

$$\tilde{\mathcal{L}} = \frac{1}{2}\frac{\partial^2}{\partial \tilde{x}^2} + \rho \sigma_{\tilde{y}} \frac{\partial^2}{\partial \tilde{x} \partial \tilde{y}} + \frac{1}{2}\sigma_{\tilde{y}}^2 \frac{\partial^2}{\partial \tilde{y}^2}, \qquad \tilde{\Omega} := (0, 1) \times (0, 1) \tag{3.17}$$

with $\sigma_{\tilde{y}} = \min\{\tau_x/\tau_y, \tau_y/\tau_x\}$ subject to the initial and boundary conditions

$$\begin{aligned} p_{(2)}(\tilde{x},\tilde{y},\tilde{t}) &= 0, \text{ for } (\tilde{x},\tilde{y}) \in \partial\tilde{\Omega}, \\ p_{(2)}(\tilde{x},\tilde{y},0) &= \delta\left(\tilde{x}-\tilde{x}_0\right)\delta\left(\tilde{y}-\tilde{y}_0\right). \end{aligned} \tag{3.18}$$

As seen above, the computational domain under $T_{(2)}$ remains the unit square. However, the diffusion coefficient in the principal $\tilde{x}$-direction is always unity, while $\sigma_{\tilde{y}} \leq 1$. The time for evaluating the final condition is scaled by either τ_x^2 or τ_y^2, while the correlation coefficient remains the same. Transforming back to the original coordinate frame is straight-forward

$$p(x,y,t) = p_{(2)}(\tilde{x}(x,y;\sigma_x,\sigma_y,\rho), \tilde{y}(x,y;\sigma_x,\sigma_y,\rho), \tilde{t}(t;\sigma_x,\sigma_y,\rho)) \cdot \frac{1}{L_x L_y}.$$

Hence, the solution to the original problem (3.4) - (3.5) can be expressed as the product

$$\begin{aligned} q(x,y,t) = {} & \exp(-\alpha x - \beta y - \gamma t)\cdot \\ & p_{(2)}(\tilde{x}(x,y;\sigma_x,\sigma_y,\rho), \tilde{y}(x,y;\sigma_x,\sigma_y,\rho), \tilde{t}(t;\sigma_x,\sigma_y,\rho)) \cdot \frac{1}{L_x L_y}. \end{aligned}$$

As a consequence of the second remark above, all of the dependency of q on (a_x, b_x, a_y, b_y) is through the normalized solution $p_{(2)}(\tilde{x},\tilde{y},\tilde{t})$. Thus, without loss of generality, we will concentrate on the solution of the normalized problem (3.16) and will drop the subscript $p_{(2)}$ from here on when discussing the solution to the normalized problem.

3.3 A Review of Standard Approaches

Before describing our method for solving the normalized problem (3.16) - (3.18), we discuss other candidate approaches and highlight their limitations in the context of the present problem.

3.3.1 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:

$$p(\tilde{x},\tilde{y},\tilde{t}) = \sum_{\nu} h_\nu \phi_\nu(\tilde{x},\tilde{y}) e^{-\lambda_\nu \tilde{t}}, \tag{3.19}$$

where h_ν is the coefficient of the eigenfunction $\phi_\nu(\tilde{x},\tilde{y})$. 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 decays to zero as $\tilde{t}$ goes to infinity (see section 6.3 of Risken [1989]).

Since we require $\phi_\nu(\tilde{x},\tilde{y})$ to be zero on the boundaries, we approximate the eigenfunction using a finite set of orthogonal basis functions satisfying boundary conditions, i.e., a finite sequence of sines,

$$\phi_\nu(\tilde{x},\tilde{y}) = \sum_{l=1}^{L}\sum_{m=1}^{M} c_{l,m,\nu} \sin(2\pi\, l\, \tilde{x}) \sin(2\pi\, m\, \tilde{y}) := \Psi(\tilde{x},\tilde{y})^T c_\nu,$$

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

$$\begin{aligned}\psi_{l,m}(\tilde{x},\tilde{y}) &= \sin\left(2\pi\, l\, \tilde{x}\right)\sin\left(2\pi\, m\, \tilde{y}\right),\\ \Psi(\tilde{x},\tilde{y}) &= (\psi_{1,1}(\tilde{x},\tilde{y}),\ldots,\psi_{L,M}(\tilde{x},\tilde{y}))^T,\\ c_\nu &= (c_{1,1,\nu},\ldots,c_{L,M,\nu})^T.\end{aligned}$$

We substitute the biorthogonal representation (3.19) into the eigenvalue problem

$$\mathcal{L}\phi_\nu = -\lambda_\nu\phi_\nu, \tag{3.20}$$

where $\mathcal{L}$ is the differential operator in the normalized Fokker-Planck equation. Applying $\mathcal{L}$ to the basis function expansion of ϕ_v and again approximating the result using the finite set of basis functions yields

$$\mathcal{L}\phi_\nu = \mathcal{L}(\Psi(\tilde{x},\tilde{y})^T c_\nu) = \mathcal{L}(\Psi(\tilde{x},\tilde{y})^T)c_\nu = (A\Psi(\tilde{x},\tilde{y}))^T c_\nu, \tag{3.21}$$

where A is a matrix analytic in $(\sigma_{\tilde{y}},\rho)$. In the last part of the equation above, we have truncated the infinite sine series expansion of $\mathcal{L}\psi_{l,m}(\tilde{x},\tilde{y})$. In the case where $\rho = 0$, A is diagonal because $\{\psi_{l,m}(\tilde{x},\tilde{y})\}$ are the eigenfunctions to $\mathcal{L}$. When $\rho \neq 0$, A is no longer diagonal and is in fact dense. Substituting the linear representation of $\mathcal{L}\phi_\nu$ into the eigenvalue problem (3.20), we arrive at the matrix eigenvalue problem

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

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

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

Because a good working approximation of $p(\tilde{x}, \tilde{y}, \tilde{t})$ requires many terms in the expansion, and because the resultant system matrix is dense, repeated computation of the eigenproblem for the density calculation is unfeasible. This is especially important in the context of the main problem: computing derivatives of p with respect to the boundary parameters.

3.3.2 Finite Difference Methods

A finite difference method which approximates the spatial derivatives in problem (3.16) - (3.18) requires the solution to a system of differential equations

$$\dot{c}(\tilde{t}) = Bc(\tilde{t}) \Rightarrow \qquad c(\tilde{t}) = \exp\left(B\tilde{t} \right) c(0), \tag{3.22}$$

which reduces to the eigenvalue decomposition of a matrix B. Here, $c(\tilde{t})$ is a vector consisting of values of the solution in (3.16) on a set of grid points over $\tilde{\Omega}$ at time $\tilde{t}$, and the product $Bc(\tilde{t})$ approximates $\mathcal{L}p(\tilde{x}, \tilde{y}, \tilde{t})$. The system matrix B is dependent on the discretization scheme used to approximate $\mathcal{L}$. Using a central-in-space scheme over a **regular** grid on $\tilde{\Omega}$ with $\Delta\tilde{x} = \Delta\tilde{y} = h$, we have

$$B = \frac{1}{2}\frac{1}{h^2} B_{\tilde{x},\tilde{x}} + \rho\sigma_{\tilde{y}} \frac{1}{4h^2} B_{\tilde{x},\tilde{y}} + \frac{1}{2}\sigma_{\tilde{y}}^2 \frac{1}{h^2} B_{\tilde{y},\tilde{y}},$$

where each of the matrices $B_{\tilde{x},\tilde{x}}, B_{\tilde{x},\tilde{y}}, B_{\tilde{y},\tilde{y}}$ approximate $\frac{\partial^2}{\partial \tilde{x}^2}, \frac{\partial^2}{\partial \tilde{x}\partial \tilde{y}}, \frac{\partial^2}{\partial \tilde{y}^2}$, respectively. The use of a regular grid with a constant h independent of the parameters is appealing here because, it allows us to construct once and store the matrices

$B_{\tilde{x},\tilde{x}}, B_{\tilde{x},\tilde{y}}, B_{\tilde{y},\tilde{y}}$, which saves valuable computational resources if we are to solve the finite difference eigenproblem (3.22) repeatedly for different parameter values $(\sigma_{\tilde{y}}, \rho)$.

Unlike the system matrix for the trigonometric expansion, the system matrix B here is sparse: each row of $(B_{\tilde{x},\tilde{x}}, B_{\tilde{x},\tilde{y}}, B_{\tilde{y},\tilde{y}})$ 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. However, the solutions generated by this methods lead to quite inaccurate finite difference approximations to the derivatives of p which, as the reader might recall, are the main object of interest in our problem.

3.3.3 Calculation of Joint Density

Following (3.6), to compute the joint density $f(x, y, a_x, b_x, a_y, b_y)$, we must take derivatives of the solution $p(\tilde{x}, \tilde{y}, \tilde{t})$ with respect to the boundary parameters (a_x, b_x, a_y, b_y). Because the approximate solutions we will consider below are functions of the parameters without explicit analytic form, $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(\tilde{x}, \tilde{y}, \tilde{t})$ must be computed numerically using a finite difference approximation. We must therefore find the solution for each of the sixteen perturbed problems

$$p(\tilde{x}, \tilde{y}, \tilde{t} \mid a_x \pm \epsilon, b_x \pm \epsilon, a_y \pm \epsilon, b_y \pm \epsilon).$$

The derivative of p with respect to the 4 boundaries is approximated as

$$\begin{aligned} &\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(\tilde{x}, \tilde{y}, \tilde{t}) \\ &\quad \approx \frac{\sum\limits_{k_1,k_2,k_3,k_4=\pm 1} c_{\{k_1,k_2,k_3,k_4\}} p(\tilde{x}, \tilde{y}, \tilde{t} \mid a_x + k_1\varepsilon, b_x + k_2\varepsilon, a_y + k_3\varepsilon, b_y + k_4\varepsilon)}{(2\varepsilon)^4} \end{aligned}$$

The finite difference approximation, however, introduces fundamental limitations for the type of approximation that can be used and the degree of accuracy admissible for this problem.

Consider an approximate solution $p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t} \mid b)$ where we have included parameter b explicitly as a simplified notation for $[a_x, b_x]$ and $[a_y, b_y]$ and we assume that the approximation approaches the true solution as the discretization resolution $k \to \infty$. In general, the truncation error can be represented as

$$p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}|b) - p(\tilde{x}, \tilde{y}, \tilde{t}|b) = \left(\frac{1}{k}\right)^{\alpha} \mathrm{F}_{reg}(b) + \left(\frac{1}{k}\right)^{\beta} \mathrm{F}_{irreg}(b) + \varepsilon_{mach} F_{round}(b), \tag{3.23}$$

for some $\alpha, \beta > 0$; $(1/k)^{\alpha}\mathrm{F}_{reg}(b)$ is the regular part of the truncation error, which is a smooth function of b; $(1/k)^{\beta}\mathrm{F}_{irreg}(b)$ is the irregular part of the truncation error, which is not smooth in b; and $\varepsilon_{mach}\mathrm{F}_{round}(b)$ is the effect of round-off errors with $\varepsilon_{mach} \sim 10^{-16}$ denoting the machine epsilon for IEEE double precision system.

Note that when expressed using the chain rule, both $\dfrac{\partial}{\partial a_x}$ and $\dfrac{\partial}{\partial b_x}$ contain $\dfrac{\partial}{\partial b}$. As a result, $\dfrac{\partial^2}{\partial a_x \partial b_x}$ leads to $\dfrac{\partial^2}{\partial b^2}$. Although in the discussion below, for simplicity, we only illustrate the numerical differentiation on the first derivative, keep in mind that it is the second derivative that is more relevant in the calculation of $\dfrac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(\tilde{x}, \tilde{y}, \tilde{t})$.

The coefficient, $\mathrm{F}_{reg}(b)$, of the regular part of truncation error is a smooth function of b with derivative $= O(1)$. The coefficient, $\mathrm{F}_{round}(b)$, in the effect of round-off errors, behaves virtually like a random variable, discontinuous in b. For the irregular part of truncation error, the coefficient $\mathrm{F}_{irreg}(b)$ can be thought of in general as continuous in b but not smooth in b where the derivative has discontinuities. Linear interpolation applied to accommodate starting position or ending position not falling on a grid point, for example, introduces this kind of irregular truncation error. Based on the expression we wrote out above for the finite difference solution, applying the numerical differentiation with step ε yields:

$$\begin{aligned}\frac{p^{(k)}(\tilde{x},\tilde{y},\tilde{t}\big|\, b+\varepsilon)-p^{(k)}(\tilde{x},\tilde{y},\tilde{t}\big|\, b-\varepsilon)}{2\varepsilon} &= \\ &\frac{\partial}{\partial b}p(\tilde{x},\tilde{y},\tilde{t}\big|\, b)+O(\varepsilon^2)+\left(\frac{1}{k}\right)^{\alpha}\frac{\mathrm{F}_{reg}(b+\varepsilon)-\mathrm{F}_{reg}(b-\varepsilon)}{2\varepsilon} \\ &\qquad +\left(\frac{1}{k}\right)^{\beta}\frac{\mathrm{F}_{irreg}(b+\varepsilon)-\mathrm{F}_{irreg}(b-\varepsilon)}{2\varepsilon} \\ &\qquad\qquad +\varepsilon_{mach}\frac{\mathrm{F}_{round}(b+\varepsilon)-\mathrm{F}_{round}(b-\varepsilon)}{2\varepsilon}.\end{aligned}$$

In the equation above, as the step in the numerical differentiation is refined, the first line of the RHS is well behaved, converging to the true value $\frac{\partial}{\partial b}p(\tilde{x},\tilde{y},\tilde{t}\mid b)$ as $\varepsilon\to 0$. The second line of RHS, however, is problematic. As $\varepsilon\to 0$, the contribution from round-off error blows up to infinity

$$\varepsilon_{mach}\frac{\mathrm{F}_{round}(b+\varepsilon)-\mathrm{F}_{round}(b-\varepsilon)}{2\varepsilon}=O\left(\frac{\varepsilon_{mach}}{\varepsilon}\right)\longrightarrow\infty\quad\text{as }\varepsilon\to 0$$

For a fourth-order derivative, the error contribution of finite machine precision becomes $O(\varepsilon_{mach}/\varepsilon^4)$ and a step size as big as $\varepsilon\sim 10^{-4}$ produces errors $O(1)$. This is catastrophic in instances where true values of the fourth order derivative (the joint density) are smaller than 1.

3.3.4 Boundary derivatives with finite difference and irregular truncation errors

The contribution from the irregular part of truncation error can also be problematic if the finite difference approximation of the derivative of $F_{irreg}(b)$ is $O(k^{\beta}/\varepsilon)$. Indeed, this is a property of the finite difference method because of the linear interpolation necessary for function arguments not on grid points. Referring to the notation in (3.23), and setting $k := 1/h$, the interpolation error introduced by a finite difference scheme at position b for the boundary parameters using step h has the general form of

$$\text{Interpolation error } = O(h^2)(1 - \text{remainder}(b/h, 1))\text{remainder}(b/h, 1)$$

The coefficient part $\mathrm{F}_{irreg}(b) = (1 - \text{remainder}(b/h, 1))\text{remainder}(b/h, 1)$ is continuous in b but not differentiable. Its first derivative has the behavior of

$$\frac{\partial}{\partial b}\mathrm{F}_{irreg}(b) = \frac{1}{h}\left(1 - 2\text{remainder}(b/h, 1)\right).$$

The contribution from the irregular part of truncation error is

$$h^2 \frac{\mathrm{F}_{irreg}(b+\varepsilon) - \mathrm{F}_{irreg}(b-\varepsilon)}{2\varepsilon} = O\left(\frac{h^2}{\max(\varepsilon, h)}\right)$$

In the second order numerical differentiation, however, the contribution from the irregular part of truncation error behaves like

$$h^2 \frac{\mathrm{F}_{irreg}(b+\varepsilon) - 2\mathrm{F}_{irreg}(b) + \mathrm{F}_{irreg}(b-\varepsilon)}{\varepsilon^2} = O\left(\frac{h^2}{\max(\varepsilon^2, h^2)}\right)$$

The interplay between h and ε limits the size ε we can use to perform density calculations for a fixed step size h.

Hence, while finite difference approximations lead to sparse systems that can be solved relatively fast, they impose a natural limit on the quality of the derivative. On the other hand, eigenvalue expansions do not suffer from having irregular truncation errors (leading to more accurate derivatives), but the system matrix is dense and therefore the problems are slow to solve.

3.4 A Novel Semidiscrete Galerkin Method

We propose a semidiscrete Galerkin (continuous in time, discrete in space) solution to the general diffusion problem (3.9). The method relies on an analytic approximation for the solution $p(\tilde{x},\tilde{y},\tilde{t})$ for small time $\tilde{t}$ and a basic convergence estimate from approximation theory for parabolic problems.

Our numerical method produces a functional representation of the approximate solution which (*i*) imposes a computational burden comparable to or better than that of the finite difference method and (*ii*) is infinitely differentiable with respect to the boundary parameters, allowing us to perform the crucial numerical differentiation needed to obtain the density of interest.

As described in Section 3.3.1, the solution to the model problem (3.16) - (3.18) has the eigenfunction expansion

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

where h_ν is the coefficient of $\phi_\nu(\tilde{x},\tilde{y})$ in the eigenfunction expansion of $p(\tilde{x},\tilde{y},0)$. Proceeding with the standard Galerkin approach, we propose a solution $p^{(k)}(\tilde{x},\tilde{y},\tilde{t})$ of similar form

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

where the basis functions $\psi_i(\tilde{x},\tilde{y})$ satisfy the boundary conditions. We also require that all first- and second-order derivatives of $\psi_i(\tilde{x},\tilde{y})$ are in $L_2(\tilde{\Omega})$, i.e. $\psi_k(\tilde{x},\tilde{y}) \in W_2^2(\tilde{\Omega})$. This will allow us to take derivatives of the approximate solution with respect to the boundaries for the problem. Finally, we designate the set of basis functions as

$$S_k := \{\psi_1(\tilde{x},\tilde{y}), \ldots, \psi_k(\tilde{x},\tilde{y})\}.$$

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

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

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:

$$\int_\Omega R_e(k)\psi_i(\tilde{x}, \tilde{y}) d\tilde{x}\, d\tilde{y} = 0, \quad \int_\Omega R_0(k)\psi_i(\tilde{x}, \tilde{y}) d\tilde{x}\, d\tilde{y} = 0, \quad i = 0, \dots k, \tag{3.24}$$

which is equivalent to the weak formulation of the heat problem

$$\begin{aligned} \left\langle \partial_t p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}), \psi_i \right\rangle &= \left\langle \tilde{\mathcal{L}} p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}), \psi_i \right\rangle, \\ \left\langle p^{(k)}(\tilde{x}, \tilde{y}, 0), \psi_i \right\rangle &= \langle p(\tilde{x}, \tilde{y}, 0), \psi_i \rangle, \end{aligned}$$

where $< \cdot, \cdot >$ is the usual inner product in $L_2(\tilde{\Omega})$. The orthogonality conditions (3.24) lead to the system of equations

$$M\dot{\mathbf{c}}(\tilde{t}) = S\mathbf{c}(\tilde{t}), \tag{3.25}$$

$$M\mathbf{c}(0) = \mathbf{p}(0), \tag{3.26}$$

where M is the mass matrix, S is the stiffness matrix, and each component of $\mathbf{p}(0)$ is the vector projection of $p(\tilde{x}, \tilde{y}, 0)$ onto the corresponding vector set in S_k.

The matrices M and S, and the vector $\mathbf{p(0)}$ are constructed as:

$$\begin{aligned}
[M]_{ij} &= \int_{\tilde{\Omega}} \psi_i \psi_j d\tilde{x}\, d\tilde{y}, \\
[S]_{ij} &= -\frac{1}{2} \int_{\tilde{\Omega}} \left(\frac{\partial}{\partial \tilde{x}} \psi_i(\tilde{x}, \tilde{y}) \right) \left(\frac{\partial}{\partial \tilde{x}} \psi_j(\tilde{x}, \tilde{y}) \right) d\tilde{x}\, d\tilde{y} \\
&\quad - \rho \sigma_{\tilde{y}} \int_{\tilde{\Omega}} \left(\frac{\partial}{\partial \tilde{x}} \psi_i(\tilde{x}, \tilde{y}) \right) \left(\frac{\partial}{\partial \tilde{y}} \psi_j(\tilde{x}, \tilde{y}) \right) d\tilde{x}\, d\tilde{y} \\
&\quad - \frac{1}{2} \sigma_{\tilde{y}}^2 \int_{\tilde{\Omega}} \left(\frac{\partial}{\partial \tilde{y}} \psi_i(\tilde{x}, \tilde{y}) \right) \left(\frac{\partial}{\partial \tilde{y}} \psi_j(\tilde{x}, \tilde{y}) \right) d\tilde{x}\, d\tilde{y}, \\
[\mathbf{p}(0)]_i &= \int_{\tilde{\Omega}} p(\tilde{x}, \tilde{y}, 0) \psi_i(\tilde{x}, \tilde{y}) d\tilde{x}\, d\tilde{y}.
\end{aligned}$$

The entries of S_{ij} are computed with integration by parts to enforce the boundary conditions. Then, the semidiscrete Galerkin approximation becomes

$$p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}) = \boldsymbol{\psi}(\tilde{x}, \tilde{y})^T \exp\left(M^{-1} S\, \tilde{t} \right) \mathbf{c}(0),$$

with $\boldsymbol{\psi}(\tilde{x}, \tilde{y}) = (\psi_0(\tilde{x}, \tilde{y}), \ldots, \psi_k(\tilde{x}, \tilde{y}))^T$. Since both M and S are infinitely differentiable with respect (a_x, b_x, a_y, b_y) (each entry in the matrices has this property, which can be seen in the defining expressions above), the system matrix $M^{-1}S$ is infinitely differentiable with respect to the parameters (a_x, b_x, a_y, b_y) as well. As a consequence, the matrix exponential

$$\exp\left(M^{-1} S\, \tilde{t} \right) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(M^{-1} S \right)^n \tilde{t}^n$$

is also infinitely differentiable with respect to the boundary parameters. Hence, we can safely apply $\frac{\partial}{\partial a_x \partial b_x \partial a_y \partial b_y}$ to the Galerkin approximation $p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t})$ as long as the derivatives of the vectors $\boldsymbol{\psi}(\tilde{x}, \tilde{y})^T$ and $\mathbf{c}(0)$ exist. This is ensured as long as each $\psi_i(\tilde{x}, \tilde{y})$ is differentiable with respect to the boundary parameters, and therefore the approximation of $f(x, y, a_x, b_x, a_y, b_y)$ based on $p^{(k)}$ exists.

In general, both M and S are dense. Thus, the computational complexity of the method is determined by the number of eigenfunctions needed to represent the solution and, in general, need not be more efficient than a finite difference method. However, unlike the trigonometric expansion approach in Section 3.3.1, the Galerkin method presented here admits a much more general family of basis functions which need not be separable. As a result, if chosen to accommodate the correlation ρ in the differential operator, the basis family can yield systems that are small enough to outperform the finite difference approach yet still accurate enough to be useful. Below, we describe our choice of basis functions which satisfy this criteria.

3.4.1 Basis Family

We motivate the construction of the basis functions by considering the fundamental solution for the unbounded problem (3.16). We choose the family of basis functions $S_k = \{\psi_i(\tilde{x}, \tilde{y}),\ 0 \leq i \leq k\}$

$$\begin{aligned}\psi_i(\tilde{x}, \tilde{y}) = &\frac{1}{2\pi\tilde{\sigma}^2\sqrt{1-\tilde{\rho}^2}} \\ &\times \exp\left\{-\frac{((\tilde{x}-\tilde{x}_i)^2 - 2\tilde{\rho}(\tilde{x}-\tilde{x}_i)(\tilde{y}-\tilde{y}_i) + (\tilde{y}-\tilde{y}_i)^2)}{2(1-\tilde{\rho}^2)\tilde{\sigma}^2}\right\} \\ &\times \tilde{x}\,(1-\tilde{x})\ \tilde{y}(1-\tilde{y})\end{aligned} \tag{3.27}$$

for some parameters $(\tilde{\rho}, \tilde{\sigma})$ and a collection of nodes $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=0}^{k}$ which form a grid over $\tilde{\Omega}$.

This grid is determined by the choice of kernel parameters $(\tilde{\rho}, \tilde{\sigma})$ with a scaling parameter l, and it is defined in the following way. For a set of $(\tilde{\rho}, \tilde{\sigma}, l)$, shift the coordinate system so that $(\tilde{x}'_0, \tilde{y}'_0) = (1/2, 1/2)$. In the shifted coordinate system, generate a rectangular grid, uniform in each of x- and y-directions, over the square

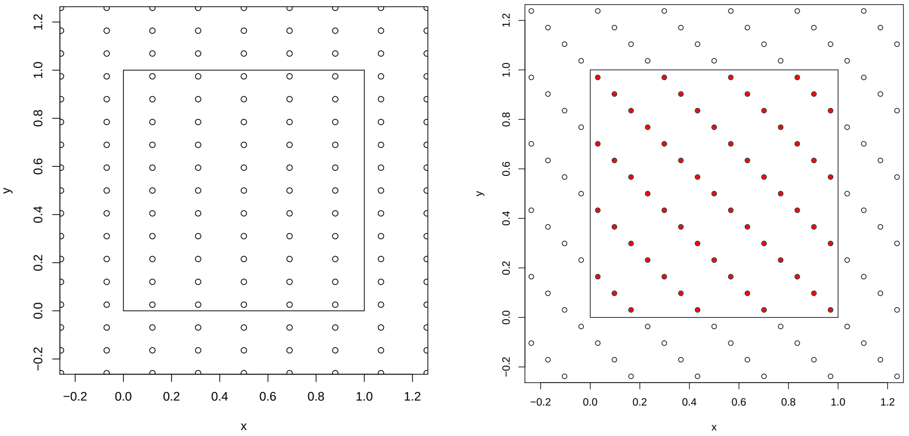

Figure 3.1: A sample grid design for $l = 1$, $\sigma = 0.3$ and $\rho = 0.6$. The left panel corresponds to the initial grid $\{(x'_j, y'_j)\}_{j=0}^{k'}$ over Ω (solid black square). The right panel depicts the rotated initial grid. The set of final node points $\{(x_i, y_i)\}_{i=0}^{k}$ is contained within Ω and is denoted by the red solid points.

region $[1/2 - 1/\sqrt{2}, 1/2 + 1/\sqrt{2}] \times [1/2 - 1/\sqrt{2}, 1/2 + 1/\sqrt{2}]$. The grid size in the x-direction is set to $l\tilde{\sigma}(1+\tilde{\rho})$ and the grid size in the y-direction is $l\tilde{\sigma}(1-\tilde{\rho})$. The square region is selected to ensure that after a $\pi/4$ rotation around $(1/2, 1/2)$, the rotated grid should cover the square $[0,1] \times [0,1]$ (see left panel of Figure 3.1).

The grid is comprised of all nodes within $\tilde{\Omega}$:

$$\{(\tilde{x}_i, \tilde{y}_i)\}_{i=0}^{k} = \left\{(\tilde{x}_j, \tilde{y}_j) | (\tilde{x}_j, \tilde{y}_j) \in \tilde{\Omega}, j = 0, \ldots, k'\right\},$$

(see right panel of Figure 3.1). It should be noted that the level sets of the heat kernel for the basis functions are ellipses with major and minor axes aligned with the node points. Further, there is more resolution (ie. the node layout is denser) along the direction corresponding to the smaller standard deviation of the basis heat kernel in the principal coordinate frame. Finally, for a given l, nodes in either principal direction are separated by l standard deviations of the basis heat kernel. This layout naturally takes into account some degree of correlation $\tilde{\rho}$ so that the small-time solution, having level sets similar to those of the kernels, can be better resolved. Essentially, the collection $\{\psi_i(\tilde{x}, \tilde{y}|\tilde{\rho}, \sigma)\}_{i=0}^{k}$ is composed of fundamental solutions to a heat diffusion problem tuned by σ and $\tilde{\rho}$, tampered such that their support is on $\tilde{\Omega}$, zero on the boundaries, still smooth, and inheriting the correlation structure of the fundamental solution to the problem. For the purposes of this paper, we found that keeping fixed $l = 1$ with a moderate $\tilde{\rho}$ ($|\tilde{\rho}| \sim 0.8$) yields reasonable results. It is important, however, that $\text{sign}(\tilde{\rho}) = \text{sign}(\rho)$ so that the problematic, narrow component of the small-time solution (which can be seen in the minor axis of the contours of the small-time solution in the right panel of Figure 3.2) can be resolved.

3.4.2 Small-time Solution

The initial condition for our problem requires that the basis family used for the Galerkin method resolve very high-wave number terms necessary to approximate a $\delta-$like function. In order to avoid this and thereby reduce the computational load for our method, we develop an analytic, small-time solution to the problem. In addition to attenuating high-frequency terms by analytically evolving the process forward in time for a short period, we also reduce the overall error of our method, as will be demonstrated in Section 3.4.3.

The small-time solution is derived by enforcing only one boundary condition, at the boundary that has the largest influence on the solution. We start with the fundamental solution $G_\rho(\tilde{x}, \tilde{y}, \tilde{t}|\tilde{x}_0, \tilde{y}_0)$ for the unbounded problem in (3.16), 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 $\tilde{t}_\epsilon$ such that $G_\rho\left(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon \,|\tilde{x}_0, \tilde{y}_0\right)$ is numerically zero on the rest of the boundaries of $\tilde{\Omega}$. We select the largest among all time instances satisfying this condition. This is done in the following way.

1. Scale and rotate the coordinate axes by the transformation $T_{(3)}$

$$T_{(3)} : \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{1-\rho}} & -\frac{1}{\sigma_{\tilde{y}}\sqrt{1-\rho}} \\ \frac{1}{\sqrt{1+\rho}} & \frac{1}{\sigma_{\tilde{y}}\sqrt{1+\rho}} \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}, \tag{3.28}$$

$$p(\xi, \eta, \tilde{t}) = \sigma_{\tilde{y}}\sqrt{1-\rho^2} \cdot p(\tilde{x}, \tilde{y}, \tilde{t}),$$

so that the problem obeys the standard diffusion equation

$$\frac{\partial}{\partial \tilde{t}} p(\xi, \eta, \tilde{t}) = \frac{1}{2}\frac{\partial^2}{\partial^2 \xi} p(\xi, \eta, \tilde{t}) + \frac{1}{2}\frac{\partial^2}{\partial^2 \eta} p(\xi, \eta, \tilde{t})$$

on a computational domain now transformed to a parallelogram $\Omega^{(3)}$ (see left panel of Figure 3.2). The transformed initial condition will be denoted as (ξ_0, η_0).

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

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

We define the **distance** between $G(\xi, \eta, \tilde{t}|\xi_0, \eta_0)$ and any of the linear boundaries of $\Omega^{(3)}$ as the shortest distance between (ξ_0, η_0) to each of the boundary segments (blue line segments in the left panel of Figure 3.2). There are four such distances d_1, d_2, d_3, d_4; and assume that they are listed in increasing magnitude. Note that the perpendicular intersections may not fall within the line segment when the four line segments form a very skewed diamond. At the line segment where we enforce the boundary condition, the perpendicular intersection from (ξ_0, η_0) onto the boundary is always contained in the line segment. This boundary is nearest to (ξ_0, η_0), corresponds to the shortest one of four distances.

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

3. Reflect the point $(\xi_0, \eta_0) \to (\xi_0', \eta_0')$ about the closest boundary. The image function $G(\xi, \eta, \tilde{t}_\epsilon | \xi_0', \eta_0')$ satisfies the diffusion equation and is equal to $G(\xi, \eta, \tilde{t}_\epsilon | \xi_0, \eta_0)$ on the closest boundary. Further, $G(\xi, \eta, \tilde{t}_\epsilon | \xi_0', \eta_0')$ takes on values less than 10^{-15} on all other boundaries, because it is outside of $\Omega^{(3)}$. For this same reason, the system of images numerically satisfies the initial condition for the problem.

   The *small-time* solution p_ϵ is the difference of the two images

   $$p_\epsilon(\xi, \eta, \tilde{t}) := G(\xi, \eta, \tilde{t} | \xi_0, \eta_0) - G(\xi, \eta, \tilde{t} | \xi_0', \eta_0'),$$

   satisfying all of the boundaries numerically and also satisfying the governing diffusion equation for $\tilde{t} \leq \tilde{t}_\epsilon$, as illustrated in the left panel of Figure 3.2. Performing a change of variables $T_{(3)}^{-1} : (\xi, \eta) \to (\tilde{x}, \tilde{y})$ produces the small-time solution $p_\epsilon(\tilde{x}, \tilde{y}, \tilde{t})$ in the normalized diffusion problem frame

   $$p_\epsilon(\tilde{x}, \tilde{y}, \tilde{t}) = G_\rho(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_0, \tilde{y}_0) - G_\rho(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_0', \tilde{y}_0'), \qquad (3.29)$$

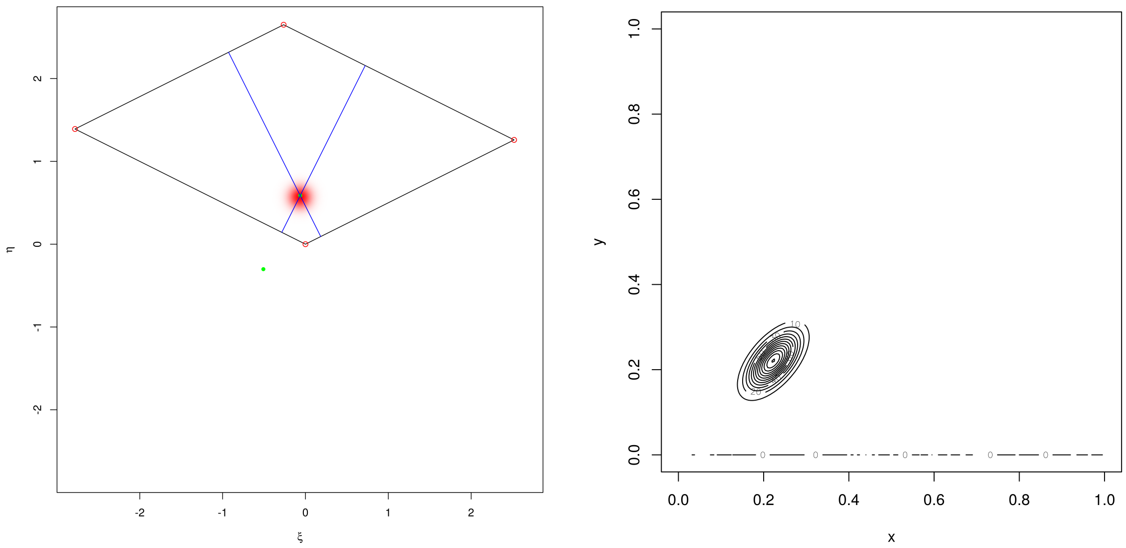

Figure 3.2: An example of the small-time solution $p(\xi, \eta, t_\epsilon)$ on the transformed domain $\tilde{\Omega}$ with $\tau_x = \tau_y = 1$ and $\rho = 0.6$. **Left**: The shaded red region is a heatmap of the small-time solution in the transformed coordinate frame, while the blue line segments represent the distance between the boundaries and the initial condition coordinate. The green point outside of the computational domain is the center of the reflected image (ξ_0', η_0') about the closest boundary. **Right**: The small-time solution transformed back to the original coordinate system. Here, the contours denote the level-sets for the function. They very closely approximate the level sets for the fundamental solution of the unbounded problem (3.9)

where the function $G_\rho(\cdot)$ is the shifted fundamental solution

$$G_\rho(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_0, \tilde{y}_0) =$$
$$\frac{1}{2\pi\ \tilde{t}\sigma_{\tilde{y}}\sqrt{1-\rho^2}} \exp\left(-\frac{\left[(\tilde{x}-\tilde{x}_0)^2 \sigma_{\tilde{y}}^2 + (\tilde{y}-\tilde{y}_0)^2 - 2\rho(\tilde{x}-\tilde{x}_0)(\tilde{y}-\tilde{y}_0)\sigma_{\tilde{y}} \right]}{2\ \tilde{t}\, \sigma_{\tilde{y}}^2 (1-\rho^2)} \right)$$

as illustrated in the right panel of Figure 3.2.

Using Theorem 5.E of Zeidler [1995], we can solve for $p(\tilde{x}, \tilde{y}, \tilde{t})$ by considering the smooth $p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)$ as an initial condition and evolving it forward in time by

$\tilde{t} - \tilde{t}_\epsilon$. This replaces initial condition in (3.26) with

$$\begin{aligned} M\mathbf{c}(\tilde{t}_\epsilon) &= \mathbf{p}(\tilde{t}_\epsilon), \\ [\mathbf{p}(\tilde{t}_\epsilon)]_i &= \int_{\tilde{\Omega}} p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)\psi_i(\tilde{x}, \tilde{y}) d\tilde{x}\, d\tilde{y}. \end{aligned} \tag{3.30}$$

Intuitively, the bigger $\tilde{t}_\epsilon$, the smaller both errors $R_e(k)$ and $R_0(k)$ will be in approximating the solution for the initial value $p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)$. On the other hand, for large $\tilde{t}_\epsilon$, the error in the small-time approximate solution will be significant. The selection of $\tilde{t}_\epsilon = d_2/8$ is a good compromise for accommodating these two opposing forces in reducing errors.

3.4.3 Error Bound

A bound on the closeness of the approximate solution $p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t})$ to the strong solution $p(\tilde{x}, \tilde{y}, \tilde{t})$ is developed in Bramble et al. [1977]. Their result shows that the Galerkin approximation we use converges to the strong solution in $L_2(\tilde{\Omega})$, and it motivates the thrust of our numerical solution. First, we define the *error* term

$$e^{(k)}(t) = p(\tilde{x}, \tilde{y}, \tilde{t}) - p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}),$$

as well as the norm

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

for the eigenpairs (λ_j, ϕ_j) of the operator $\mathcal{L}$. The $\|\cdot\|_2$ norm used in Bramble et al. [1977] captures the variation in a given function: for any set of eigenpairs $\{(\lambda_j, \phi_j)\}$ over the domain $\tilde{\Omega}$, increasing the absolute value of λ_j with j corresponds to a function ϕ_j with increasing number of oscillations over $\tilde{\Omega}$. The $\|\cdot\|_2$ norm is distinct from the usual $L_2(\tilde{\Omega})$ norm , which we denote as $\|\cdot\|_{L_2(\tilde{\Omega})}$. As referred to in Bramble et al. [1977], functions $w \in L_2(\tilde{\Omega})$ with $\|w\|_2 < \infty$ are also in $W_2^2(\tilde{\Omega})$. Finally, if we have the condition (corresponding to equation 2.1 in Bramble et al. [1977])

$$\|p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon) - p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)\|_{L_2(\tilde{\Omega})} \leq C\, h(k)^2 \|p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)\|_2, \tag{3.31}$$

where $h(\cdot)$ is a decreasing function of $k > 0$, Theorem 2.1 in Bramble et al. [1977] applies and we have the error estimate

$$\|e^{(k)}(t)\|_{L_2(\tilde{\Omega})} \leq Ch(k)^2 \|p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)\|_2. \tag{3.32}$$

Here, the constant C and the function $h(k)$ are the same as in (3.31). The implication is that if the basis functions in S_k represent the small-time solution $p(\tilde{x}, \tilde{y}, \tilde{t}_\epsilon)$ with no error, the Galerkin solution forward in time is also without error.

We can ensure condition (3.31) is met if S_k is complete in $L_2(\tilde{\Omega})$ as k grows. The other two conditions necessary for the error bound to apply are demonstrated by Bramble et al. [1977] for the Galerkin method. Equation (3.32) can be summarized in a simple way: the *error* of the method is controlled by how much variation the small-time solution has; the *rate* of decrease of the error is controlled by how well the span of S_k represents the small-time solution compared to the variation of the initial condition as k increases. In the context of (3.32), our method, with its small-time analytic solution and choice of basis functions, is specifically tailored to minimize the error between the strong solution and its Galerkin approximation under $L_2(\tilde{\Omega})$.

3.5 Estimation

The pdf from (3.6) provides the likelihood necessary for frequentist and Bayesian inference in bivariate settings. Likelihood-free approaches, like that of Rogers and Satchell [1991] and Rogers et al. [2008], on the other hand suffer from being suboptimal when data are independent and identically distributed (i.i.d.) and cannot be easily extended to the non-i.i.d. case.

To illustrate the suboptimality of likelihood-free estimators for the i.i.d. case, we consider maximum likelihood inference from an i.i.d. sample $Z_1, \ldots, Z_n$, where $Z_i = (X_i, Y_i, m_{x,i}, M_{x,i}, m_{y,i}, M_{y,i})$, where X_i and Y_i corresponds to log returns of the two assets, $m_{x,i}$ and $m_{y,i}$ correspond to the minimum log-returns observed during the day, and $M_{x,i}$ and $M_{y,i}$ correspond to the maximum log returns. Computation of the maximum likelihood estimation (MLE) is based on an iterative derivative-free numerical algorithm to maximize the likelihood (the Nelder-Mead method; see Lagarias et al. [1998] for review and convergence properties). This is feasible even for moderate to large sample sizes, because our numerical method is specifically designed for computational efficiency for repeatedly evaluating the density function (3.6).

3.5.1 Consistency

In this section we prove that the MLE based on the Galerkin approximation $p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t})$ is consistent. To do so, we first show that the distribution on Z based on the approximate $p^{(k)}(\tilde{x}, \tilde{y}, \tilde{t})$ converges to the true distribution $F(\cdot|\theta)$. Define the approximate probability density

$$f^{(k)}(x, y, a_x, b_x, a_y, b_y) := \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}). \tag{3.33}$$

At this point we will assume that for a sufficiently large k, $f^{(k)}(z)$ is positive for all $z \in Z$ and is integrable over Z. As such, we may regard it as a proper probability density function with the cumulative probability density and probability measure over Z being

$$F^{(k)}(z|\theta) = \int_{-\infty}^{a_x} \int_{-\infty}^{a_y} \int_{-\infty}^{b_x} \int_{-\infty}^{b_y} \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{\partial^4}{\partial a'_x \partial b'_x \partial a'_y \partial b'_y} q^{(k)}(x', y', t | a'_x, b'_x, a'_y, b'_y) \; dz, \tag{3.34}$$

$$\Pr_k(A) := \int_A f^{(k)}(z)\, dz, \qquad \text{for any measurable } A \subset Z, \tag{3.35}$$

where $z = (x, y, a_x, b_x, a_y, b_y)$. We will prove that for every $z \in Z$,

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

First, we prove the Lemma

Lemma 1.

$$\lim_{k \to \infty} \int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y)\, dx\, dy = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y)\, dx\, dy.$$

Proof. Define the sets of form, for $z = (x, y, a_x, b_x, a_y, b_y)$,

$$\begin{aligned} &B(a_x, b_x, a_y, b_y) = \\ &\quad \left\{ z' \in Z \,|\, z' \in [a_x, b_x] \times [a_y, b_y] \times [a_x, \infty) \times (-\infty, b_x] \times [a_y, \infty) \times (-\infty, b_y] \right\}. \end{aligned}$$

Elements within $B(a_x, b_x, a_y, b_y)$ are equivalent to sample paths that stay within the region $[a_x, b_x] \times [a_y, b_y]$:

$$\begin{aligned} &B(a_x, b_x, a_y, b_y) = \\ &\qquad \{\omega \in W | X_\omega(t) \in [a_x, b_x], Y_\omega(t) \in [a_y, b_y], \\ &\qquad\quad m_x \in [a_x, b_x), M_x \in (a_x, b_x], m_y \in [a_y, b_y), M_y \in (a_y, b_y]\} . \end{aligned}$$

Then

$$\begin{aligned} \Pr(B(a_x, b_x, a_y, b_y)) &= \int_{B(a_x, b_x, a_y, b_y)} f(z)\, dz \\ &= \int_{a_y}^{\infty} \int_{-\infty}^{b_y} \int_{a_x}^{\infty} \int_{-\infty}^{b_x} \int_{a_y}^{b_x} \int_{a_x}^{b_x} f(x', y', a'_x, b'_x, a'_y, b'_y) dx'\, dy'\, da'_x\, db'_x\, da'_y\, db'_y \\ &= \int_{a_x}^{b_x} \int_{a_y}^{b_y} q(x, y, a_x, b_x, a_y, b_y)\, dx\, dy, \end{aligned} \tag{3.36}$$

where the last equality employs the interpretation of q in (3.3) and we can freely change the order of integration as f is bounded for all $z \in Z$. Similarly,

$$\Pr_k(B(a_x, b_x, a_y, b_y)) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} q^{(k)}(x, y, a_x, b_x, a_y, b_y)\, dx\, dy.$$

From (3.36), the partial derivative of the probability $\Pr(B(a_x, b_x, a_y, b_y))$ is well-defined and can be written as

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y)\, dx\, dy.$$

The second-order finite difference approximation the above expression can be expressed as a linear combination of probabilities of perturbed sets $B(\cdot)$:

$$\lim_{\epsilon\to 0} \quad \frac{1}{\epsilon^4}\sum_{i=1}^{16} c(i)\Pr(B(a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon) = \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y}\Pr(B(a_x, b_x, a_y, b_y)),$$

where $c(i)$ and $k(i)$ are functions mapping index i to the corresponding coefficient in the second-order finite difference approximation $c(i) \to \{-1, 1\}$, $k_j(i) \to \{-1, 1\}$. Using Big-O notation, for a sufficiently small ϵ

$$\sum_{i=1}^{16} c(i)\Pr(B(a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon) = \epsilon^4\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y}\Pr(B(a_x, b_x, a_y, b_y)) + O(\epsilon^6; a_x, b_x, a_y, b_y). \quad (3.37)$$

The convergence result in Section 3.4.3 implies that

$$\begin{aligned}\Pr_k(B(a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon) &= \\ \int_{a_x+k_1(i)\epsilon}^{b_x+k_2(i)\epsilon}\int_{a_y+k_3(i)\epsilon}^{b_y+k_4(i)\epsilon} q^{(k)}(x, y, a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon)\,dx\,dy &\to \\ \int_{a_x+k_1(i)\epsilon}^{b_x+k_2(i)\epsilon}\int_{a_y+k_3(i)\epsilon}^{b_y+k_4(i)\epsilon} q(x, y, a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon)\,dx\,dy &= \\ \Pr(B(a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon) \text{ as } k\to\infty \text{ in } L_2(\Omega).\end{aligned}$$

Hence, for a sufficiently large k dependent on the supremum over i, and given the error estimate in (3.32), we obtain the relation

$$\begin{aligned}\Pr_k(B(a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon) &= \\ \Pr(B(a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon) \\ + O(h(k)^2; a_x+k_1(i)\epsilon, b_x+k_2(i)\epsilon, a_y+k_3(i)\epsilon, b_y+k_4(i)\epsilon)\end{aligned}$$

Note here that the dominating terms $O(h(k)^2; \cdot)$ are differentiable with respect to the boundary parameters (a_x, b_x, a_y, b_y) since q and $q^{(k)}$ have this property. Therefore, if we replace $\Pr(\cdot)$ in (3.37) with $\Pr_k(\cdot)$

$$\begin{aligned}
&\sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) = \\
&\sum_{i=1}^{16} c(i) \Pr(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) \\
&+ \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) = \\
&\epsilon^4 \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\epsilon^6; a_x, b_x, a_y, b_y) \\
&+ \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon).
\end{aligned}$$

Dividing both sides by ϵ^4 produces

$$\begin{aligned}
&\frac{1}{\epsilon^4} \sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) = \\
&\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\epsilon^2; a_x, b_x, a_y, b_y) \\
&+ \frac{1}{\epsilon^4} \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon).
\end{aligned}$$

As mentioned above $O(h(k)^2; \cdot)$ is differentiable with respect to the boundary parameters, so that the right-most term is still $O(h(k)^2)$ as $\epsilon \to 0$. Taking the limit in ϵ, we have

$$\begin{aligned}
\int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y)\, dx\, dy &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr_k(B(a_x, b_x, a_y, b_y)) \\
&= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(h(k)^2; a_x, b_x, a_y, b_y) \\
&= \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y)\, dx\, dy + O(h(k)^2; a_x, b_x, a_y, b_y).
\end{aligned}$$

Therefore, we have the desired result:

$$\lim_{k\to\infty}\int_{a_x}^{b_x}\int_{a_y}^{b_y} f^{(k)}(x,y,a_x,b_x,a_y,b_y)\,dx\,dy = \int_{a_x}^{b_x}\int_{a_y}^{b_y} f(x,y,a_x,b_x,a_y,b_y)\,dx\,dy.$$

$\square$

Lemma 2 (Convergence in distribution)**.** *For any $z \in Z$, $\lim_{k\to\infty} F^{(k)}(z|\theta) = F(z)$.*

Proof. Let

$$I^{(k)}(z) = \int_{a_x}^{x}\int_{a_y}^{y} f^{(k)}(u,v,a_x,b_x,a_y,b_y)\,du\,dv$$

and let

$$I(z) = \int_{a_x}^{x}\int_{a_y}^{y} f(u,v,a_x,b_x,a_y,b_y)\,du\,dv.$$

It is possible to show that $\lim_{k\to\infty} I^{(k)}(z) = I(z)$ as a consequence of Lemma 1 by considering some $\chi_k(z) \in S_k$ approximating the indicator $1(u \leq x, v \leq y)$ as $k \to \infty$ and setting up a triangle inequality. However, we will omit this technical detail here.

Next, we know that $I(z)$ is integrable over (a_x, b_x, a_y, b_y) as $\Pr(Z) = \mathbb{P}_W(W) = 1$. The Dominated Convergence Theorem applies, and we therefore have

$$\lim_{k\to\infty}\int_{-\infty}^{a_x}\int_{-\infty}^{b_x}\int_{-\infty}^{a_y}\int_{-\infty}^{b_y} I^{(k)}(z)da'_x db'_x da'_y db'_y = \int_{-\infty}^{a_x}\int_{-\infty}^{b_x}\int_{-\infty}^{a_y}\int_{-\infty}^{b_y} I(z)da'_x db'_x da'_y db'_y,$$

which implies the result of the Lemma. $\square$

Lemma 3. *Assuming that θ is bounded, the maximum likelihood estimator is consistent as $n \to \infty$ (number of data points) and $k \to \infty$ (number of basis elements):*

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

Proof. By Lemma 2

$$Z_k \xrightarrow{d} Z \text{ as } k \to \infty.$$

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 $Z_k \sim F_k(\theta)$,

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

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:

$$\Pr_Z\left(\left|\hat{\theta}_{n,k}(Z) - \theta\right| \geq \epsilon\right) \leq \frac{\mathrm{E}_Z\left[(\hat{\theta}_{n,k}(Z) - \theta)^2\right]}{\epsilon^2}.$$

Because the approximate likelihood function $f^{(k)}$ is continuous with respect to the data parameter z, the estimator $\hat{\theta}_{n,k}(z)$ is a continuous function with respect to z as well. Further because we have bounded $\hat{\theta}$ from below and above,

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

by the portmanteau lemma. Finally, we can show that

$$\mathrm{E}_{Z_k}\left[(\hat{\theta}_{n,k}(Z_k) - \theta)^2\right] \to 0 \text{ as } n \to \infty, \tag{3.38}$$

since the expected value of the MLE under $F(\cdot)$ tends to θ and its variance goes to 0 when $n \to \infty$. Therefore, given any $\epsilon > 0$ and $\delta > 0$, we can find a sufficiently large n and k such that

$$\Pr_{Z}\left(\left|\hat{\theta}_{n,k}(Z)-\theta\right| \geq \epsilon\right) \leq \frac{\mathrm{E}_{Z}\left[(\hat{\theta}_{n,k}(Z)-\theta)^{2}\right]}{\epsilon^{2}} < \delta$$

$\square$

3.5.2 Exploring the feasibility range for the Galerkin solution

In this section we explore the numerical limitations of the Galerkin solution in terms of its finite basis expansion and what ranges for $\tilde{t}$ can be accommodated. As described in Section 3.4.1, the semidiscrete Galerkin method that we use to compute the likelihood function requires a choice for the kernel parameters $(\tilde{\rho}, \tilde{\sigma}, l)$. Given finite restrictions on memory and compute time, $\tilde{\sigma}$ cannot be made arbitrarily small nor can $\tilde{\rho}$ be made arbitrarily close to -1 or 1 in order to accommodate all possible data. In particular, if the geometry of the problem is such that the small-time approximation does not sufficiently depart from the δ-function form of the initial condition, then there is no finite $(\tilde{\rho}, \tilde{\sigma}, l)$ combination to produce an accurate likelihood solution for arbitrarily small $\tilde{t}$ in the normalized $T_{(2)}$ frame. We illustrate this numerical limitation in Figures 3.3 - 3.5 by simulating 13 data points with $\sigma_x = 1, \sigma_y = 1, \rho = 0.95$, transforming the problem to the normalized coordinate frame, and varying $\tilde{t}$ while keeping the other parameters fixed at their data-generating values. Keeping $l = 1$ fixed, we also vary the basis-generating parameters $\tilde{\sigma}$ and $\tilde{\rho}$ to test what $\tilde{t}$ regimes can be accommodated. The finite-difference step used to derive the likelihood function is of size $\Delta = 1/32$ with $O(\Delta^2)$ order accuracy. The inner products that are needed to establish the system of equations for the method are computed using Simpson's rule on a regular

250×250 grid.

The data point in Figure 3.3 is somewhat typical of the other ones featured in Figures 3.4 - 3.5: there are no valid solutions for very small $\tilde{t}$ irrespective of basis parameter choice (usually in this regime the likelihood produced by Galerkin solution is negative and thus identified as inadmissible) and solutions converge for $\tilde{t} > 1$. The transient region between these regimes is the most problematic, particularly when the solution produces large positive values in what should be relatively low-likelihood regions; an example is the low-resolution parameter combination $(\tilde{\rho} = 0.4, \tilde{\sigma} = 0.16)$. Moreover, there is no clear sign of convergence of the solution for $\tilde{t} < 3/4$, although the combinations $(\tilde{\rho} = 0.5, \tilde{\sigma} = 0.08), (\tilde{\rho} = 0.6, \tilde{\sigma} = 0.08)$, and $(\tilde{\rho} = 0.6, \tilde{\sigma} = 0.12)$ give similar results. A feasible compromise that may operationalize the Galerkin solution in the context of maximum likelihood estimation is to use a relatively high-resolution basis expansion, such as the one corresponding to $(\tilde{\rho} = 0.6, \tilde{\sigma} = 0.08, l = 1)$ and replace inadmissible likelihood values generated for small $\tilde{t}$ with $1 \cdot 10^{-16}$. Further, this constant is also applied for likelihoods with $\tilde{t} < 0.25$, since the data points considered here clearly show that the numerical solution deteriorates in this region regardless of parameter choice and the likelihood values produced therein, even if positive, should not be used. This very rough approximation may have a variety of effects on the estimation procedure, the most obvious of which being the diverting of the convergence path of the optimization procedure regardless of algorithm used. This may be especially true in the early stages of the optimization scheme when the parameter space is being explored.

3.5.3 Simulation Study

In this section we study the maximum likelihood estimates based on the Galerkin solution with the above estimated $\tilde{t}$ cutoffs on via a simulation study.

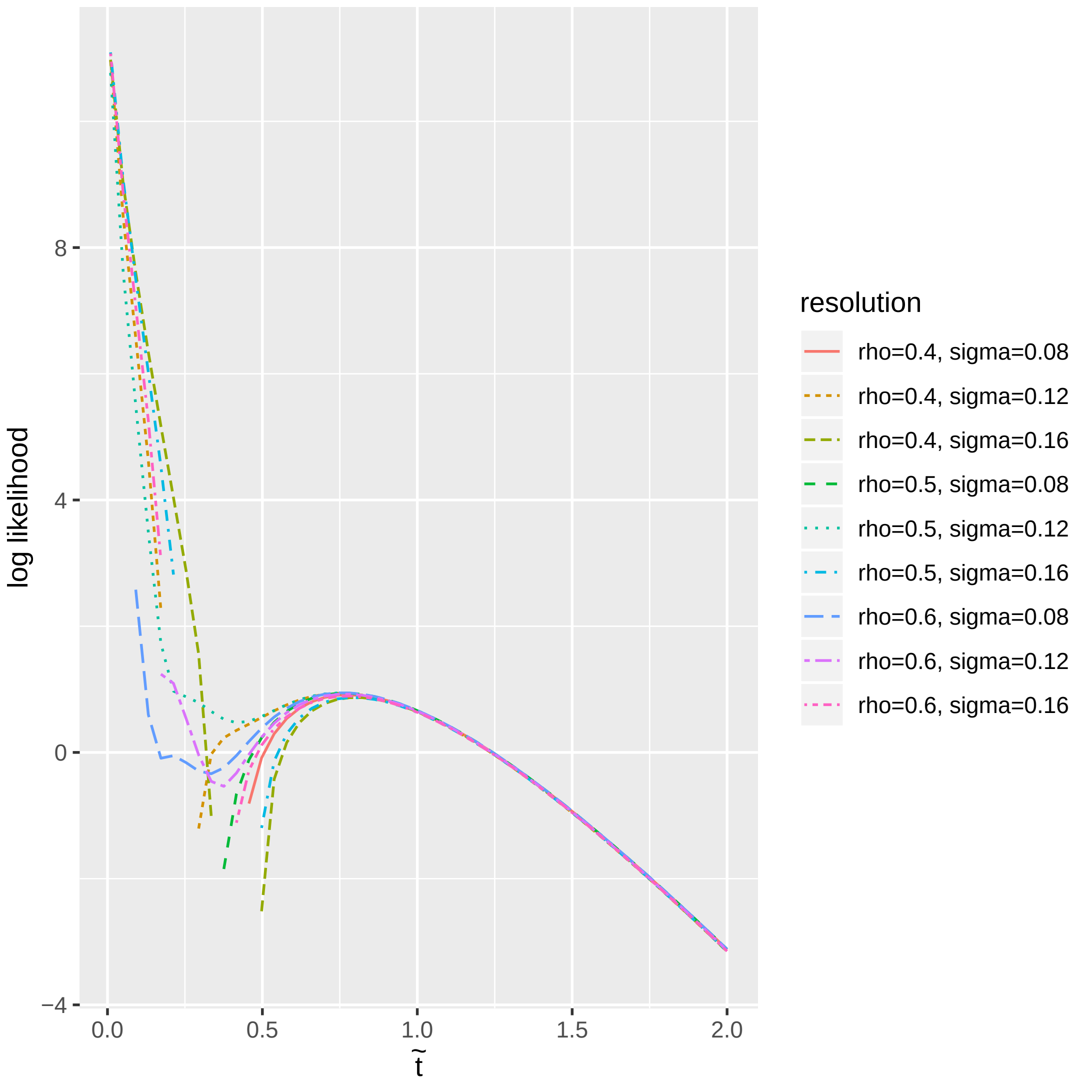

Figure 3.3: Log-likelihood as a function of $\tilde{t}$ for one of the 13 data points, computed with different $(\tilde{\rho}, \tilde{\sigma})$ combinations. The quality of the solution deteriorates for $\tilde{t} < 1/2$. For sufficiently small $\tilde{t}$, there are no valid solution values produced.

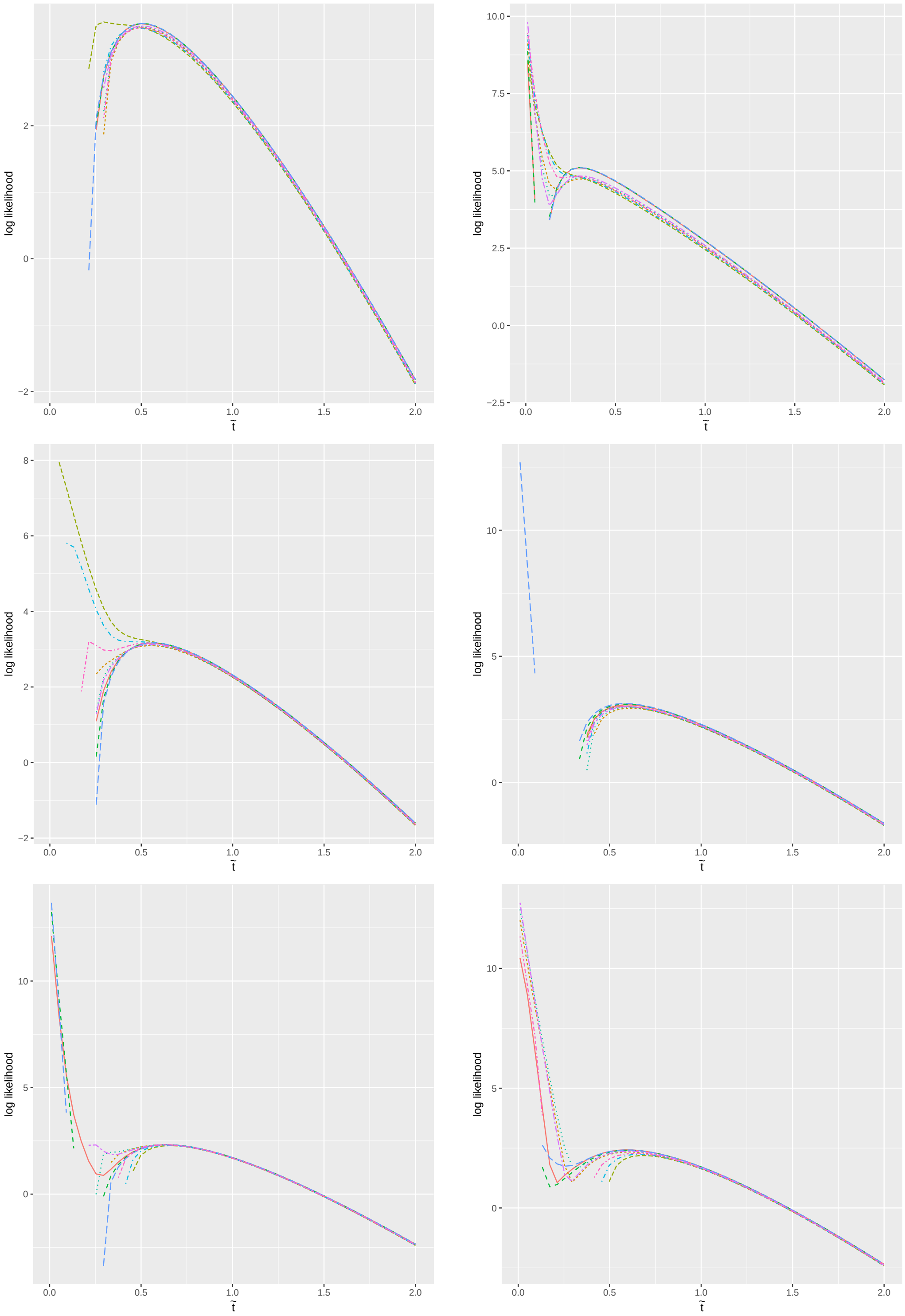

Figure 3.4: Data points 2 through 7 used to test various parameter values for basis generation. Legend is the same as in Figure 3.3.

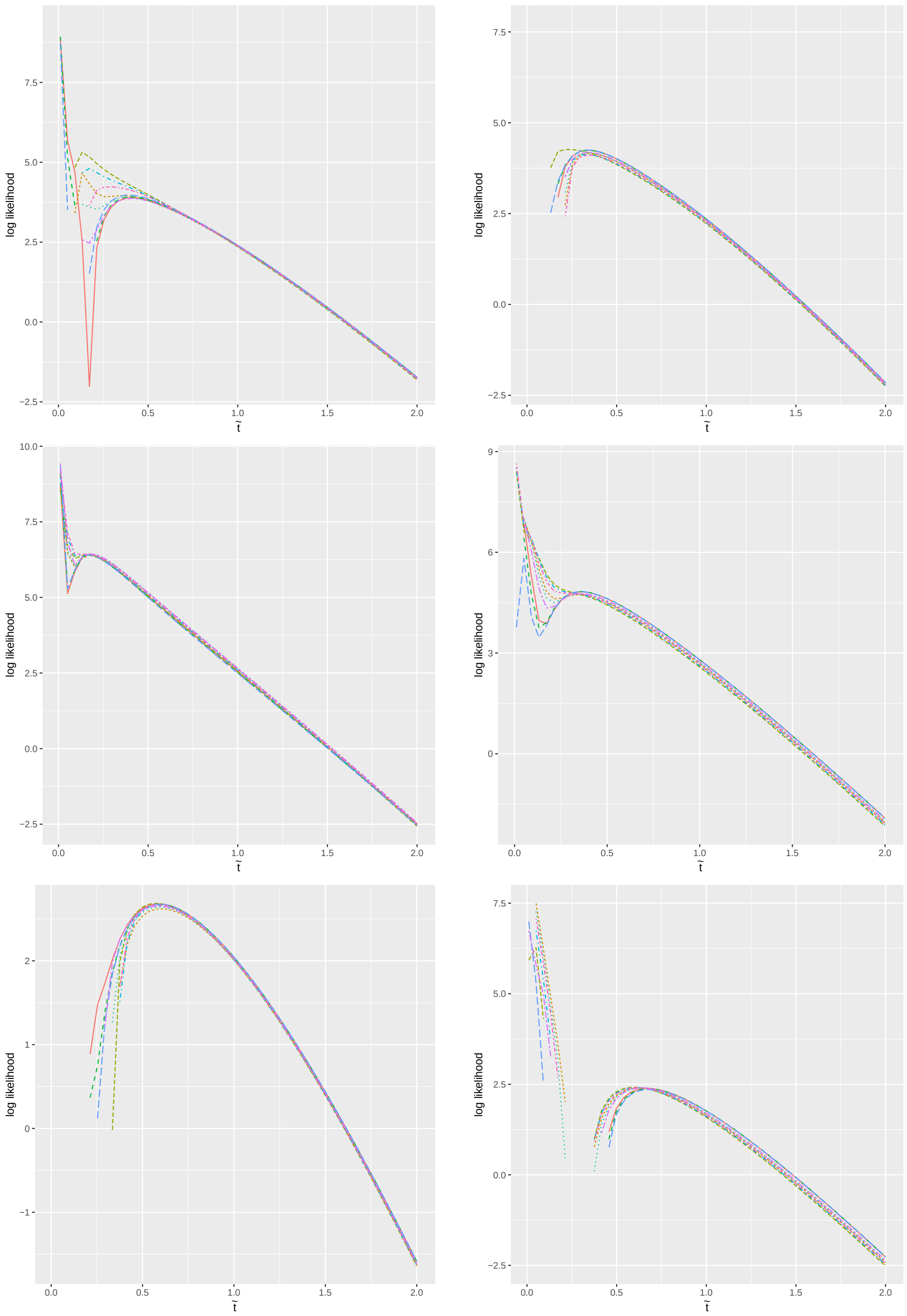

Figure 3.5: Data points 8 through 13 used to test various parameter values for basis generation. Legend is the same as in Figure 3.3.

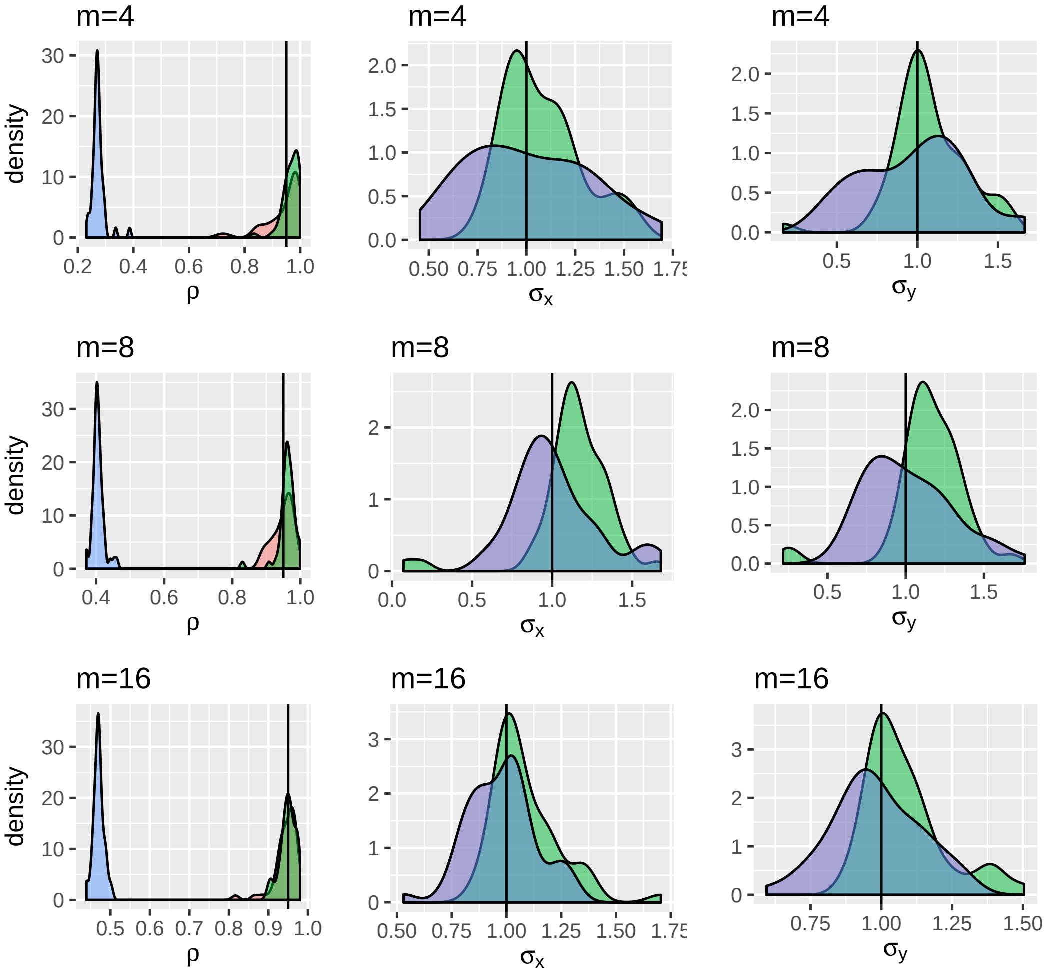

Figure 3.6: Data generated with $\rho = 0.95$. Kernel-density approximations of the repeated-sampling densities of the MLEs are shown. Samples are obtained from the Galerkin likelihood (green) and the classical likelihood (red) and the Rogers estimator (blue). The data-generating parameters are denoted with the vertical solid line.

	$\rho = 0.95$			$\rho = 0.60$			$\rho = 0.0$		
	$m = 4$	$m = 8$	$m = 16$	$m = 4$	$m = 8$	$m = 16$	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.475	1.238	1.304	0.203	0.127	0.232	0.137	0.243	0.167
$\hat{\sigma}_y$	0.593	1.040	1.088	0.111	0.120	0.260	0.189	0.171	0.107
$\hat{\rho}$	0.287	0.910	0.445	0.315	0.283	0.463	0.517	0.365	0.194

Table 3.1: Ratios of Galerkin to Gaussian MSEs for the three simulation cases.

Data is generated with zero drift via forward Euler discretization where the obtained discrete-time extremes are recorded and used as the realized extremes of the process. Simulations are generated with the same parameters

$$\mu_x = 0, \quad \mu_y = 0, \quad \sigma_x = 1, \quad \sigma_y = 1, \quad \rho = (0, 0.6, 0.95)$$

Without loss of generality, the drift parameters are assumed known, so that the MLE is comprised of the diffusion and correlation parameters: $\hat{\theta} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$. The repeated-sampling distributions of the MLEs are approximated by computing the MLE for each of $n = 50$ simulated path realizations, where each realization is divided into $m = 4, 8, 16$ equal intervals with their respective OCHL data. If we think of each path as a 6.5 hour trading day, this procedure is equivalent to estimating daily volatility and correlation for two assets using $\sim 90, 45$, and 12 min intervals throughout the trading day. Given the finite number of realizations n, we construct a kernel density estimate of the repeated-sampling distribution. Results are compared to the repeated-sampling distribution of the MLE based on the usual bivariate normal likelihood which does not take into account the boundaries, as well as the best-in-class unbiased estimator that uses OCHL data (see Rogers et al. [2008]), which we term the *Rogers estimator*.

3.6 Results and Discussion

Figure 3.6 shows the kernel density approximations of the repeated-sampling distribution for the MLE in the case of $\rho = 0.95$. The MSE ratios for the estimators are shown Table (3.1). The approximate Galerkin likelihood produces estimates with smaller MSE for ρ across all m ranges but fails to do so for the variance parameters σ_x and σ_y. However, in those cases the MSE of the Galerkin likelihood is biased by a small number of outlier estimates which are easily observed in the kernel density plots of Figure 3.6. We hypothesize that such outliers are due to the high degree of resolution in the solver necessary for the $\rho = 0.95$ case and the approximation made by replacing negative likelihood values or likelihoods computed in regions with $\tilde{t} < 0.25$ by a small numerical constant. In some cases, this can skew the optimization algorithm to regions which are not global maxima, especially in the early stages of optimization where low-likelihood regions are explored. The degree of this problem is not as great in the more mild $\rho = 0$ and $\rho = 0.60$ scenarios, as judged by the MSE ratios in Table (3.1). Yet there seem to be some cases of outliers in the estimates for the Galerkin method for these cases, as shown in Figures 3.7 and 3.8. This is more difficult to judge given the inherently greater degree of repeated sampling variability of the estimators in these scenarios. However, the combined evidence from the approximate repeated-sampling distributions, the non-uniform MSE ratios, and the break in monotonicity with respect to increasing m in the MSE of the Galerkin method (see Table (3.2)) all show that, despite providing a more powerful statistical estimator as a whole, the Galerkin likelihood developed here stands to be improved with respect to resolving likelihoods in the small-$\tilde{t}$ regions, as well as the transient region $\tilde{t}$ described above. The analytic result and asymptotic matching solution to address this need is developed in the following chapter.

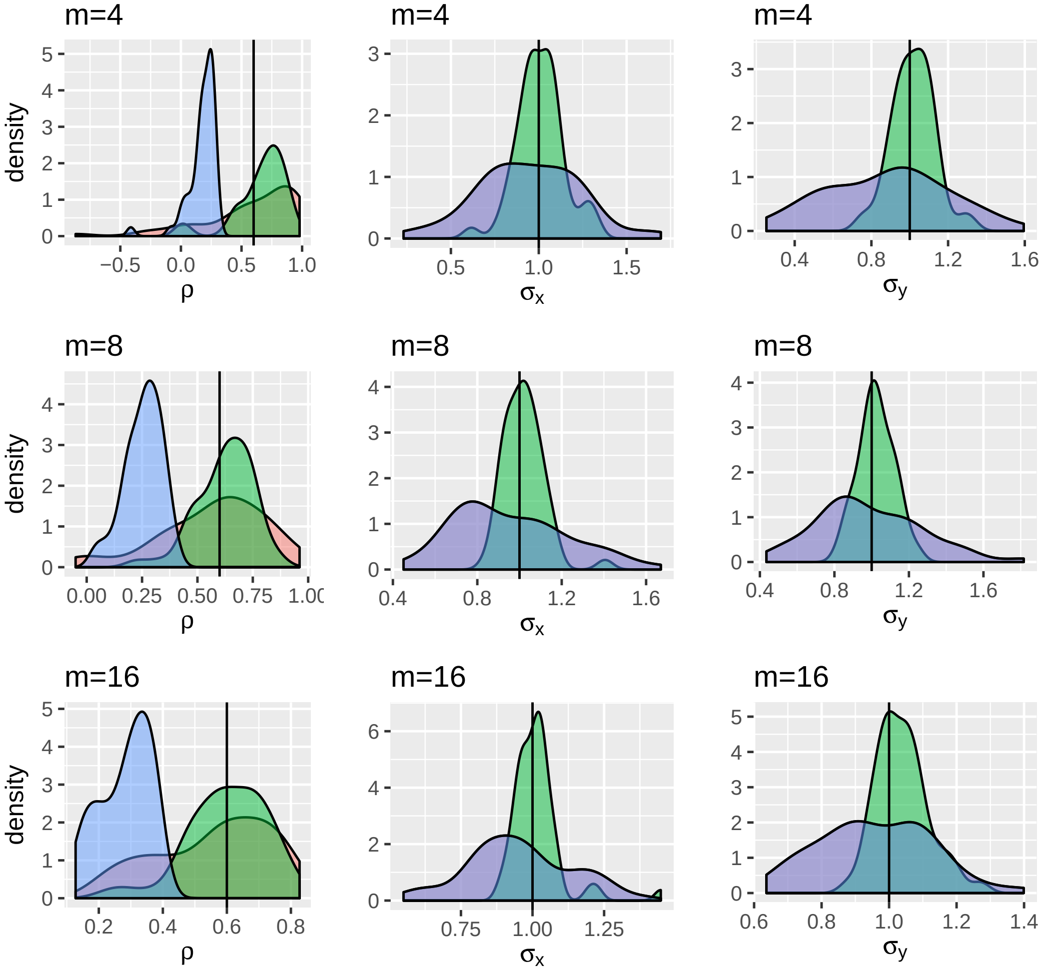

Figure 3.7: Data generated with $\rho = 0.60$. Kernel-density approximations of the repeated-sampling densities of the MLEs are shown. Samples are obtained from the Galerkin likelihood (green) and the classical likelihood (red) and the Rogers estimator (blue). The data-generating parameters are denoted with the vertical solid line.

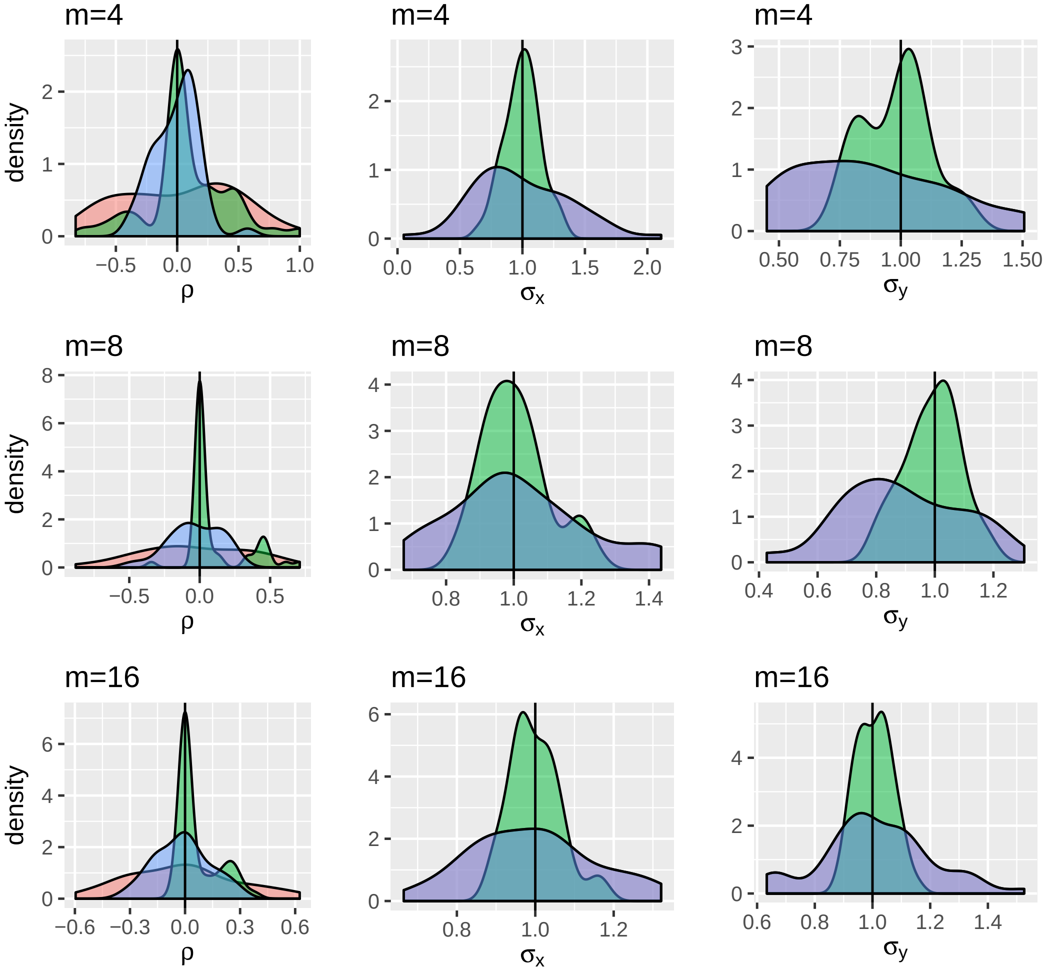

Figure 3.8: Data generated with $\rho = 0.0$. Kernel-density approximations of the repeated-sampling densities of the MLEs are shown. Samples are obtained from the Galerkin likelihood (green) and the classical likelihood (red) and the Rogers estimator (blue). The data-generating parameters are denoted with the vertical solid line.

		$\rho = 0.95$			$\rho = 0.60$	
	$m = 4$	$m = 8$	$m = 16$	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.0458	0.0825	**0.0315**	0.0175	0.00933	**0.00807**
$\hat{\sigma}_y$	0.0608	0.0772	**0.0262**	0.0124	0.00952	**0.00747**
$\hat{\rho}$	0.00120	0.000802	**0.000391**	0.0494	0.0164	**0.0148**

		$\rho = 0.0$	
	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.0458	0.0825	**0.0315**
$\hat{\sigma}_y$	0.0608	0.0772	**0.0262**
$\hat{\rho}$	0.00120	0.000802	**0.000391**

Table 3.2: MSEs for the Galerkin likelihood solution for the three simulation cases.

Chapter 4

Analytically Resolving the Transient Region of the Galerkin Solution

Chapter 3 introduced a semidiscrete (finite element) Galerkin method for solving the standardized diffusion problem (3.16). The Galerkin method is most appropriate for moderate to large times, since numerical errors are attenuated with increasing diffusion time $\tilde{t}$. Further, numerical simulations demonstrated the need to resolve the likelihood function for small $\tilde{t}$ in order to fully use the information present in OCHL data. We will call this parameter rage the *incomputable region* of the Galerkin solver. For parameters in the incomputable region, a sufficient condition to flag them as such is if the numerical derivative of the finite element solution with respect to the boundary parameters produces a negative value. However, this is not a *necessary* condition, and in practice this is an insidious problem: parameter combinations which should be attributed with a very small likelihoods may be given values orders of magnitude higher than they should and thus bias any inferential procedure (see Figure 3.3 in Chapter 3) for an illustration). In this chapter, we develop an analytic solution applicable in a small-$\tilde{t}$ region. In addition to having its own well-defined criteria for appropriate use (namely an upper bound on $\tilde{t}$ as well as an upper bound on $\tilde{\sigma}$). In this chapter we also develop an analytic matching solution that bridges the small-$\tilde{t}$ solution and the Galerkin solution across the transient $\tilde{t}$ region.

4.1 A small-time solution to the PDE, revisited

As part of the overall Galerkin method, Chapter 3 introduced a small-time approximation for the normalized diffusion problem (3.16) via the method of images (see Section 3.4.2). By reflecting the fundamental solution (i.e. the solution to the governing PDE without the boundary conditions) about the closest boundary and picking a sufficiently small $\tilde{t}_\epsilon$, the sum of images function (3.29)

$$p_\epsilon(\tilde{x}, \tilde{y}, \tilde{t}) = G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_0, \tilde{y}_0) - G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}'_0, \tilde{y}'_0) \tag{4.1}$$

satisfies the initial condition, the governing PDE, and the boundary conditions. However, because the analytic dependence of this small-time approximation on the boundaries is only through the location parameters $(\tilde{x}'_0, \tilde{y}'_0)$ of the single reflected image, differentiating with respect to all four boundaries yields a uniform zero value in computing the transition density $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(\tilde{x}, \tilde{y}, \tilde{t})$.

4.1.1 Uniqueness and Symmetry Condition

The insufficiency of the previous small-time solution suggests extending the system of images by performing more than a single reflection. If there exists an image whose location is the result of at least one reflection about each of the four boundaries, then this image is guaranteed to have a non-trivial contribution to the density $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(\tilde{x}, \tilde{y}, \tilde{t})$. An immediate problem with the extension, however, is the uniqueness of the resultant approximate density function. To illustrate the problem, consider the two systems of images constructed by carrying out the series of reflections

$$R_1 := \{1, 2, 3, 4\}, \qquad\qquad R_2 := \{2, 3, 4, 1\}, \tag{4.2}$$

where $\{r_i, r_j, r_k, r_l\}$ denotes the set of images generated by the following steps

1. r_i : Reflect the initial condition $(\tilde{x}_0, \tilde{y}_0)$ about boundary r_i to produce coordinate $(\tilde{x}_{r_i}, \tilde{y}_{r_i})$. The system of images consists of locations

   $$\{(\tilde{x}_0, \tilde{y}_0), (\tilde{x}_{r_i}, \tilde{y}_{r_i})\}$$

   with signs $\{1, -1\}$.

2. r_j : Reflect each of $\{(\tilde{x}_0, \tilde{y}_0), (\tilde{x}_{r_i}, \tilde{y}_{r_i})\}$ about boundary r_j and add to existing set of images to produce

   $$\left\{(\tilde{x}_0, \tilde{y}_0), (\tilde{x}_{r_i}, \tilde{y}_{r_i}), (\tilde{x}_{r_j}, \tilde{y}_{r_j}), (\tilde{x}_{r_j,r_i}, \tilde{y}_{r_j,r_i})\right\}$$

   with signs $\{1, -1, -1, 1\}$.

3. r_k : Reflect each of $\left\{(\tilde{x}_0, \tilde{y}_0), (\tilde{x}_{r_i}, \tilde{y}_{r_i}), (\tilde{x}_{r_j}, \tilde{y}_{r_j}), (\tilde{x}_{r_j,r_i}, \tilde{y}_{r_j,r_i})\right\}$ about boundary r_k and add to existing set of images to produce

   $$\Big\{(\tilde{x}_0, \tilde{y}_0), (\tilde{x}_{r_i}, \tilde{y}_{r_i}), (\tilde{x}_{r_j}, \tilde{y}_{r_j}), (\tilde{x}_{r_j,r_i}, \tilde{y}_{r_j,r_i}),$$
   $$(\tilde{x}_{r_k}, \tilde{y}_{r_k}), (\tilde{x}_{r_k,r_i}, \tilde{y}_{r_k,r_i}), (\tilde{x}_{r_k,r_j}, \tilde{y}_{r_k,r_j}), (\tilde{x}_{r_k,r_j,r_i}, \tilde{y}_{r_k,r_j,r_i})\Big\}$$

with signs $\{1, -1, -1, 1, -1, 1, 1, -1\}$.

4. r_l : Reflect each of the existing image locations about boundary r_l and add to the existing set of images.

Here, boundary 1 corresponds to $\tilde{y} = 0$, boundary 2 to $\tilde{x} = 1$, boundary 3 to $\tilde{y} = 1$, and boundary 4 to $\tilde{x} = 0$. Given the set $\{r_i, r_j, r_k, r_l\}$, we re-define p_ϵ as the the sum of all $J = 16$ images

$$p_\epsilon(\tilde{x}, \tilde{y}, \tilde{t}) = \sum_{j=1}^{J} (-1)^{n(j)} G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_{(j)}, \tilde{y}_{(j)}),$$

where $n(j)$ is the number of reflections performed to produce image j, and $(\tilde{x}_{(j)}, \tilde{y}_{(j)})$ is the j^{th} image location in the sequence at the end of Step 4 above. The newly defined p_ϵ analytically satisfies the boundary condition at r_l (since it is the last reflection), and it also satisfies the PDE. Once again we can choose a sufficiently small $\tilde{t}_\epsilon$ such that the boundary conditions at r_i, r_j, and r_k hold numerically as well.

Referring back to the example reflection sets in (4.2), although not identical, the two solutions to the PDE problem corresponding to R_1 and R_2 have minimal relative differences. From a numerical implementation standpoint, there is only near machine-ε difference between them. However, the two solutions have a single image, $(\tilde{x}_{1,2,3,4}, \tilde{y}_{1,2,3,4})$ and $(\tilde{x}_{2,3,4,1}, \tilde{y}_{2,3,4,1})$ respectively, whose location is a function of all four boundaries, and this single image entirely defines the corresponding density function $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p_\epsilon$. Because the location parameters of these images are different, the joint densities of $(x_0, y_0, x, y, a_x, b_x, a_y, b_y)$ derived from the two PDE solutions are consequently very different as well. Uniqueness of the solution using the method of images is achieved by performing an infinite number of reflections, which is not always possible (see Section 4.3 below).

A condition weaker than uniqueness which nonetheless restricts the solution space is the symmetry obeyed by the problem. We consider the transformation

$$\begin{aligned} x^{new} &= (a_x + b_x) - x^{old}, \\ y^{new} &= (a_y + b_y) - y^{old}. \end{aligned} \tag{4.3}$$

The PDF solution to the initial-boundary problem of the diffusion equation and joint density of $(x_0, y_0, x, y, a_x, b_x, a_y, b_y)$ are invariant with respect to this transformation. Under the normalized problem, when $(\tilde{x}_0, \tilde{y}_0) = (1/2,\ 1/2)$ and $\rho = 0$, the proposed system of images must map to itself under the corresponding coordinate transformation. Further, if we require that the system of images contain the fewest possible elements, then the unique system of images is the union of the sets of reflections

$$\{2, 4, 1, 3\} \cup \{2, 4, 3, 1\} \cup \{4, 2, 1, 3\} \cup \{4, 2, 3, 1\}\,. \tag{4.4}$$

When $(\tilde{x}_0, \tilde{y}_0) = (1/2, 1/2)$ and $\rho = 0$, the set of images from (4.4) is closed under the transformation as desired:

$$\begin{aligned} \{2, 4, 1, 3\} &\to \{4, 2, 3, 1\}, \\ \{2, 4, 3, 1\} &\to \{4, 2, 1, 3\}, \\ \{4, 2, 1, 3\} &\to \{2, 4, 3, 1\}, \\ \{4, 2, 3, 1\} &\to \{2, 4, 1, 3\}. \end{aligned}$$

This is illustrated in Figure 4.1. Removing duplicate member images of (4.4) and summing over the remaining $J^* \leq 64$, we define the new small-time solution

$$p_\epsilon(\tilde{x}, \tilde{y}, \tilde{t}) = \sum_{j=1}^{J^*} (-1)^{n(j)} G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_{(j)}, \tilde{y}_{(j)}). \tag{4.5}$$

We can control the relative error of the approximation by setting the minimum absolute value of p_ϵ at the boundaries and solving for $\tilde{t}_\epsilon$.

4.2 Calculation of the Joint Density

Out of the J^* images in (4.5), exactly four have location parameters dependent on all of the boundary parameters (a_x, b_x, a_y, b_y). Hence, the density calculation with the small-time solution becomes

$$\frac{\partial^4 p_\epsilon(\tilde{x}, \tilde{y}, \tilde{t})}{\partial a_x \partial b_x \partial a_y \partial b_y} = \sum_{j'=1}^{4} \frac{\partial^4 G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_{(j')}, \tilde{y}_{(j')})}{\partial a_x \partial b_x \partial a_y \partial b_y}. \tag{4.6}$$

One approach to compute the derivatives in (4.6) is to numerically perturb the boundary parameters (a_x, b_x, a_y, b_y) and use a finite difference approximation directly on the sum (4.5). However, since the small-time solution generally requires the use of a small time $\tilde{t} \leq \tilde{t}_\epsilon \ll 1$, numerical underflow makes direct application of finite difference hopeless. We can, however, leverage the very tractable analytic form of the Gaussian kernel $G(\cdot)$ in (3.30) by defining the following factors to more easily express the higher-order derivatives applied to (4.6)

$$\mathcal{C}(\tilde{t}, \sigma_{\tilde{y}}, \rho) := -\frac{1}{2\,\tilde{t}\,\sigma_{\tilde{y}}^2 (1-\rho^2)}, \tag{4.7}$$

$$\mathcal{P}_j(\tilde{x}, \tilde{y} | \tilde{x}_{(j)}, \tilde{y}_{(j)}) := \left(\tilde{x} - \tilde{x}_{(j)}\right)^2 \sigma_{\tilde{y}}^2 + \left(\tilde{y} - \tilde{y}_{(j)}\right)^2 - 2\rho(\tilde{x} - \tilde{x}_{(j)})(\tilde{y} - \tilde{y}_{(j)})\sigma_{\tilde{y}}. \tag{4.8}$$

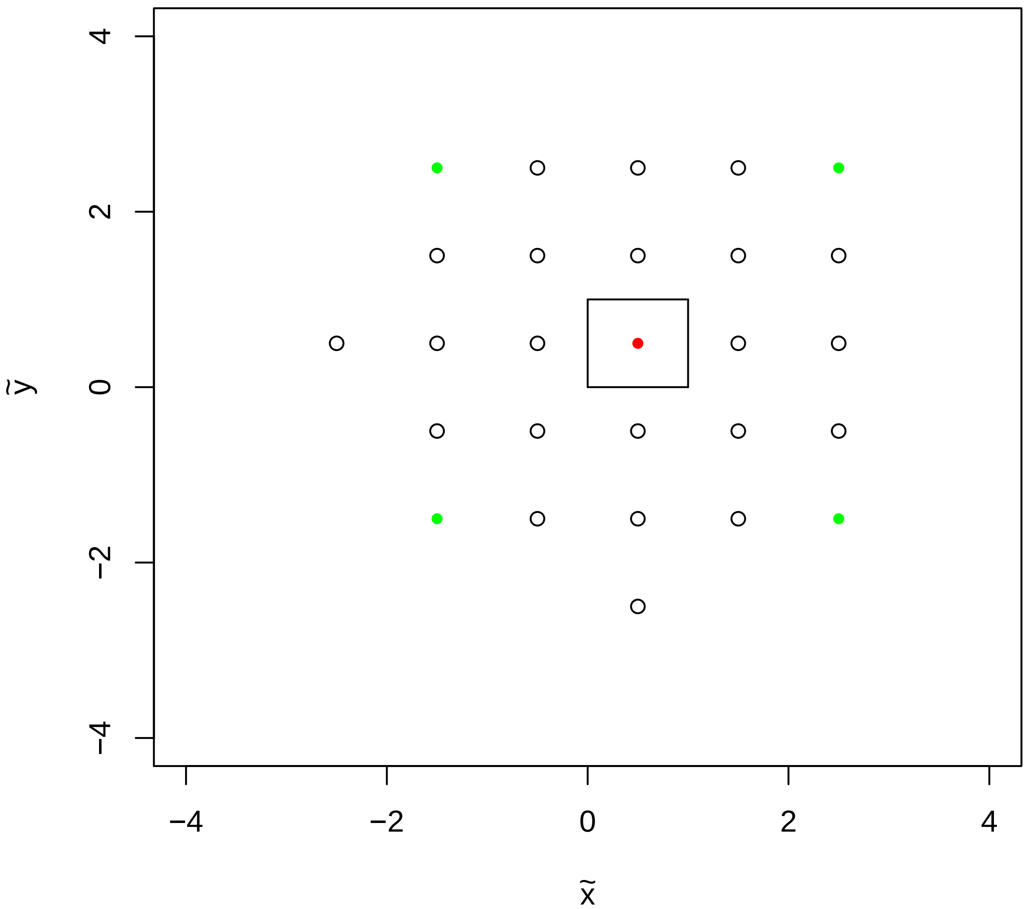

Figure 4.1: The finite system of images resultant from the sequence of reflections in the set (4.4), for $(\tilde{x}_0, \tilde{y}_0) = (1/2, 1/2)$ and $\rho = 0$. All points outside $\tilde{\Omega}$ are image position resulting from the reflections, where the solid green points have positions dependent on $(\tilde{x}_0, \tilde{y}_0)$ *as well as all of the boundaries*, and only they contribute to the likelihood solution as given in (4.6). The green colored points are symmetric about the initial condition with respect to horizontal and vertical reflections centered on $(\tilde{x}_0, \tilde{y}_0)$ and hence are closed under the transformation (4.3). Moreover, this configuration is unique under the symmetry and minimal number of images conditions.

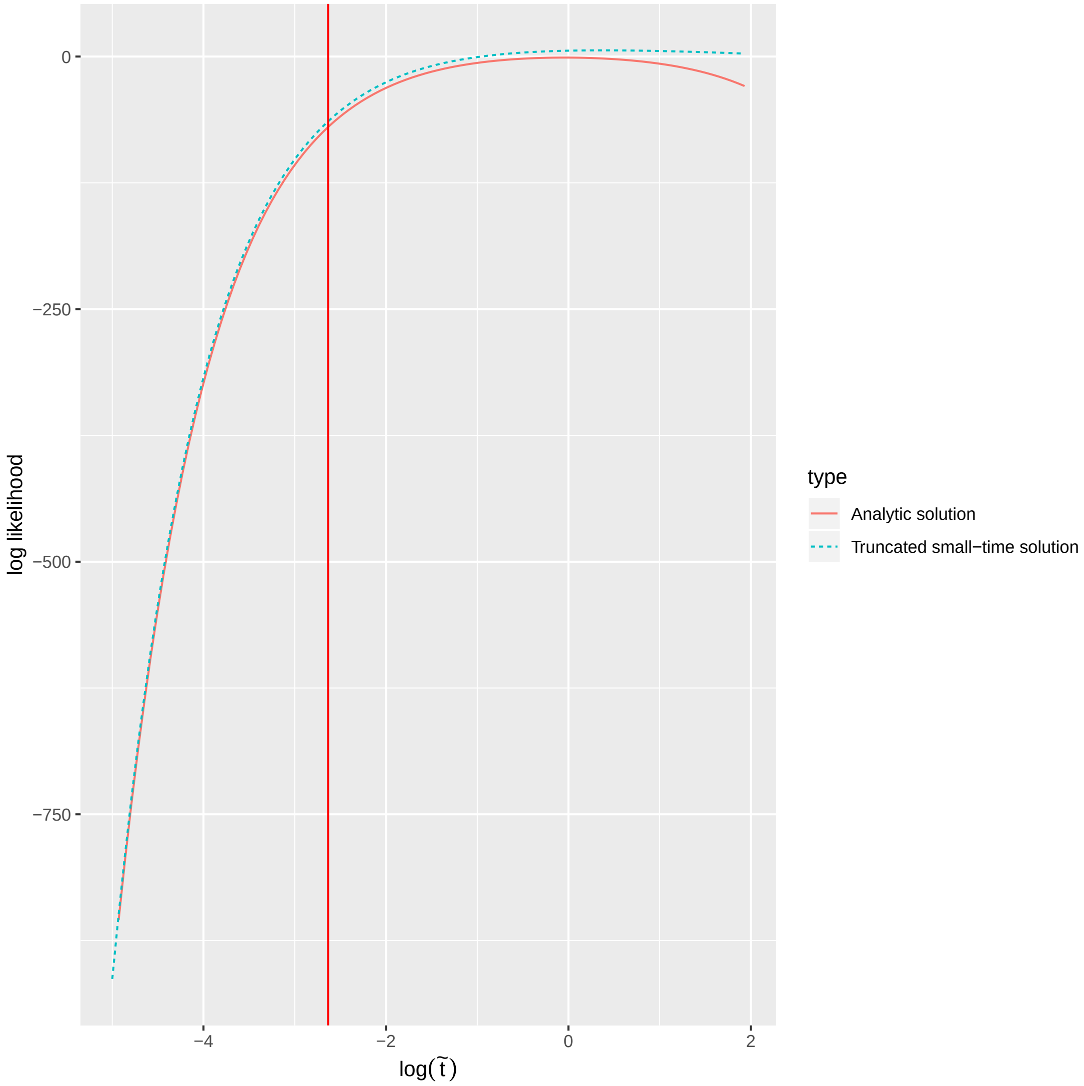

Figure 4.2: Log-likelihood profile comparison between small-time truncated solution (as given in (4.11)) and the true analytic solution for $\rho = 0$. The time up to which the untruncated small-time solution is valid is shown with the solid, vertical red line. We see that in this example the truncated approximation deteriorates for $\mathcal{O}(t) = 1$, well after the technical bound set by t^*.

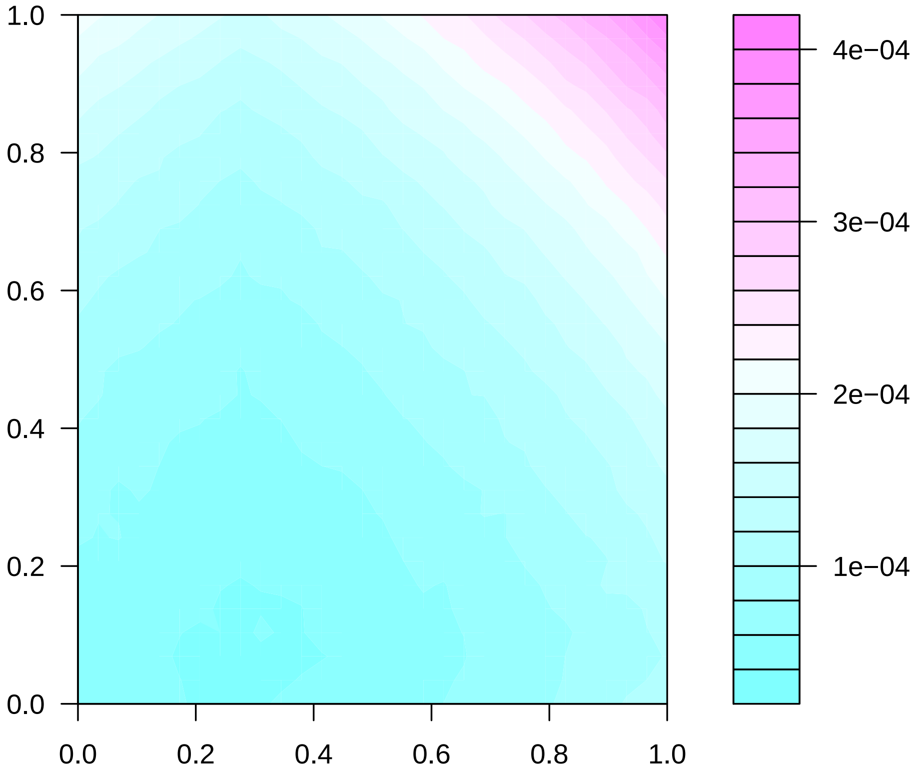

Figure 4.3: Relative error for the likelihood problem comparing the untruncated solution to the truncated small-time approximation in (4.11) on a 30×30 regular grid over Ω.

The two key observations here are that $\mathcal{P}_j(\cdot)$ is independent of $\tilde{t}$, meaning $\mathcal{C}(\cdot)$ carries the $\tilde{t}$ dependency in differentiation and that only $\mathcal{P}_j$ is dependent on the boundary parameters through $\tilde{x}_{(j)}$ and $\tilde{y}_{(j)}$. Without loss of generality, we can express the derivatives:

$$
\begin{aligned}
&\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_{(j)}, \tilde{y}_{(j)}) = \qquad (4.9)\\
&\qquad G \cdot \mathcal{C}^4 \cdot \left(\partial_{a_x} \partial_{b_x} \partial_{a_y} \partial_{b_y} \right) \mathcal{P} \\
&\qquad + G \cdot \mathcal{C}^3 \cdot \Big(\partial^2_{a_x\, b_x} \partial_{a_y} \partial_{b_y} + \partial^2_{a_x\, a_y} \partial_{b_x} \partial_{b_y} + \partial^2_{a_x\, b_y} \partial_{b_x} \partial_{a_y} + \\
&\qquad\qquad + \partial^2_{b_x\, a_y} \partial_{a_x} \partial_{b_y} + \partial^2_{b_x\, b_y} \partial_{a_x} \partial_{a_y} \partial^2_{a_y\, b_y} \partial_{a_x} \partial_{b_x} \Big) \mathcal{P} \\
&\qquad + G \cdot \mathcal{C}^2 \cdot \Big(\partial^3_{a_x\, b_x\, a_y} \partial_{b_y} + \partial^2_{a_x\, b_x} \partial^2_{a_y\, b_y} + \partial^3_{a_x\, b_x\, b_y} \partial_{a_y} + \\
&\qquad\qquad + \partial^3_{a_x\, a_y\, b_y} \partial_{b_x} + \partial^2_{a_x\, a_y} \partial^2_{b_y\, b_x} + \partial^3_{b_x\, a_y\, b_y} \partial_{a_x} + \partial^2_{b_x\, a_y} \partial_{a_x} \partial_{b_y} \Big) \mathcal{P} \\
&\qquad + G \cdot \mathcal{C} \cdot \partial^4_{a_x\, b_x\, a_y\, b_y} P, \qquad (4.10)
\end{aligned}
$$

where $\partial_x G := \partial G / \partial x$, $\partial^2_{x\,y} G := \partial^2 G / \partial x \partial y$, $\partial_x \partial_y G := \partial G / \partial x \cdot \partial G / \partial y$, etc. With (4.10), we can express each of the derivatives in (4.6) as the sum of derivatives of the polynomial $\mathcal{P}_j$ with respect to the boundary parameters. In particular, the derivatives in (4.10) avoid the previous numerical underflow problems and lend themselves to finite difference approximation even at high orders.

Thinking of $\tilde{t}$ as variable allows us to further simplify (4.10). Since $\mathcal{C} = \mathcal{O}(1/\tilde{t})$, all three terms $G \cdot \mathcal{C}^3, G \cdot \mathcal{C}^2$, and $G \cdot \mathcal{C}$ are $o\left(G \cdot \mathcal{C}^4 \cdot \left(\partial_{a_x} \partial_{b_x} \partial_{a_y} \partial_{b_y} \right) \mathcal{P} \right)$, so that the $G \cdot \mathcal{C}^4$ order term in (4.10) dominates the others for sufficiently small $\tilde{t}$. This truncation, when appropriate, is useful for reducing computational complexity *and* it guarantees that the numerical implementation of the derivatives produces positive values for all $\tilde{t}$ as long as $\left(\partial_{a_x} \partial_{b_x} \partial_{a_y} \partial_{b_y} \right) \mathcal{P}$ is positive. Truncating at $\mathcal{C}^4$, the analytic likelihood approximation given by the sum of images is of the form:

$$\frac{\partial^4 p_\epsilon(\tilde{x},\tilde{y},\tilde{t})}{\partial a_x \partial b_x \partial a_y \partial b_y} \approx \sum_{j'=1}^{4} G(\tilde{x},\tilde{y},\tilde{t}|\tilde{x}_{(j')},\tilde{y}_{(j')}) \cdot \mathcal{C}^4 \cdot \left(\partial_{a_x}\partial_{b_x}\partial_{a_y}\partial_{b_y}\right)\mathcal{P}_{j'}. \tag{4.11}$$

The above approximate solution is close to the untruncated (i.e. the one that includes all the term $\mathcal{C}^4, \mathcal{C}^3, \mathcal{C}^2$, and $\mathcal{C}^1$ term solution) small-time solution for $\tilde{t} \le \tilde{t}_\epsilon$. Extending beyond $\tilde{t}_\epsilon$ produces a positive function dominated by the hyperbolic term $\mathcal{C}^4$. Figure 4.2 shows a profile comparison of the log-likelihood with respect to $\tilde{t}$ for a representative sample configuration and $\rho = 0$. The truncated approximate solution tracks well with the true analytic solution up to $\mathcal{O}(\tilde{t}) = 1$. Applying a first-order finite difference approximation with a finite step $\Delta = 10^{-5}$ to compute the polynomial derivatives produces relative errors with a maximum of order $\mathcal{O}(10^{-4})$ on a 30×30 grid shown in Figure 4.3.

4.3 Existence of valid systems of images

For $\rho = 0$, the above procedure is guaranteed to produce an admissible set of images regardless of the initial condition and $\sigma_{\tilde{y}}$ used. However, this is not the case for a general combination of ρ and $\sigma_{\tilde{y}}$. Consider Figure 4.4, which features an initial condition $(\tilde{x}_0, \tilde{y}_0) = (0.1, 0.3)$ with $\rho = 0.9$. We can see that applying the above procedure produces a system with images within the computational domain, violating the initial condition. However, below we prove that for any $(\tilde{x}_0, \tilde{y}_0)$ and ρ combination, we can find a sufficiently small $\sigma_{\tilde{y}}$ such that a series of reflections used for p_ϵ does not violate the problem.

Lemma 4. *Given* $\rho > 0$, *and* $(\tilde{x}_0, \tilde{y}_0) \in \tilde{\Omega}$, *there exists* $0 < \sigma_{\tilde{y}} \le 1$ *such that* $(\tilde{x}_0^{(j)}, \tilde{y}_0^{(j)}) \notin \tilde{\Omega}, \forall\, j \in \{1, \ldots, J\}$, *where the collection of image locations* $\{1, \ldots, J\}$ *is the result of a finite set of reflections about the boundaries of* $\tilde{\Omega}$.

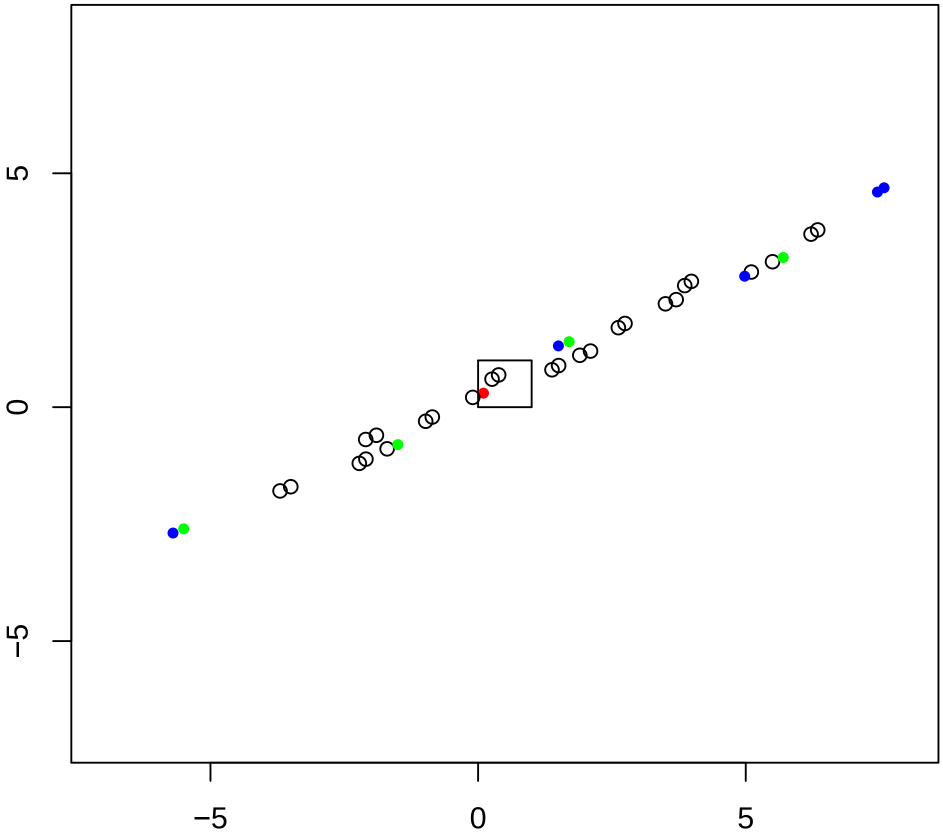

Figure 4.4: When $\rho = 0.9$, the system of images generated by the set of reflections in (4.4) and $(\tilde{x}_0, \tilde{y}_0) = (0.1, 0.3)$ now violates the initial condition as at least one of the images falls within $\tilde{\Omega}$.

Proof. Given that in the normalized problem $\tilde{\Omega}$ is a unit square whose lower-left corner is centered on the origin, applying the transformation $T_{(3)}$ in (3.28) produces a characteristic geometry for the initial initial-boundary problem that is illustrated in Figure 4.5. Corners a and c are obtuse, while b and d are acute. Further, lines $(1, 3)$ and $(2, 4)$ are each parallel and of the same length, with lines 2 and 4 being longer. This can be observed from the coordinates of the four corners

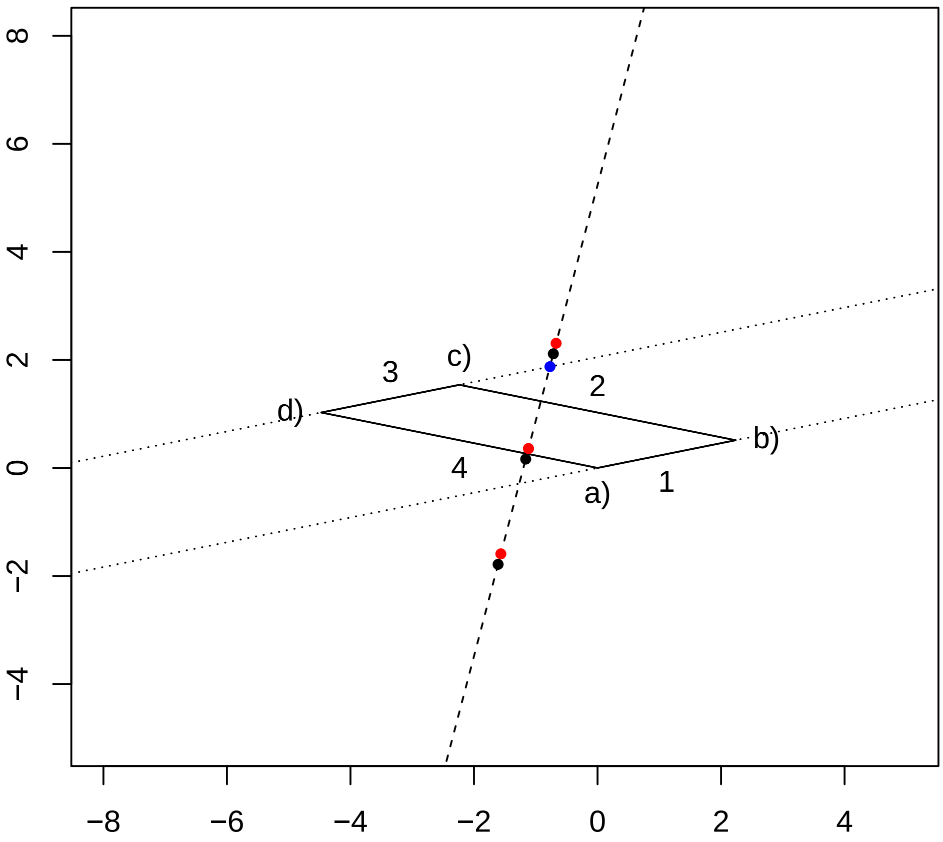

Figure 4.5: Geometry characteristic of the transformed problem when $\rho > 0$. The thick dashed line defines the axis along which images resultant from reflecting about boundaries 2 and 4 fall. The blue point denotes the intersection of the axis and the extension of boundary 3. When reflect with respect to boundary 1, for example, some of the images shown will fall into region Ω.

defining $\tilde{\Omega}$ in the transformed topology:

$$
\begin{aligned}
a &= (0, \quad 0), & b &= \frac{\sqrt{2}}{2}\left(\frac{1}{\sqrt{1-\rho}}, \quad \frac{1}{\sqrt{1+\rho}}\right), \\
c &= \frac{\sqrt{2}}{2}\left(\frac{1-1/\sigma_{\tilde{y}}}{\sqrt{1-\rho}}, \quad \frac{1+1/\sigma_{\tilde{y}}}{\sqrt{1+\rho}}\right), & d &= \frac{\sqrt{2}}{2}\left(\frac{-1/\sigma_{\tilde{y}}}{\sqrt{1-\rho}}, \quad \frac{1/\sigma_{\tilde{y}}}{\sqrt{1+\rho}}\right), \\
IC &= \frac{\sqrt{2}}{2}\left(\frac{x_0-y_0/\sigma_{\tilde{y}}}{\sqrt{1-\rho}}, \quad \frac{x_0+y_0/\sigma_{\tilde{y}}}{\sqrt{1-\rho}}\right)
\end{aligned}
$$

Now consider placing the images $j = 1, \ldots, J$ by performing a finite number of alternating reflections about lines 4 and 2. This places images along a finite segment of a line running through the initial condition position and orthogonal to lines 2 and 4 (dashed line and black/red dots in Figure 4.5). If the number of reflections is kept constant and $\sigma_{\tilde{y}}$ is sufficiently small, all thus placed images are in the interior region formed by extending lines 1 and 3. This is proved from the following observations:

1. The slopes of both lines 2 and 4 are equal to $-\frac{\sqrt{1+\rho}}{\sqrt{1-\rho}}$ and are therefore *independent of the choice* of $\sigma_{\tilde{y}}$. Hence, shrinking $\sigma_{\tilde{y}}$ moves corners d and c along lines 4 and 2, respectively, away from the origin and thereby increasing the interior region formed by extending lines 1 and 3.

2. From observation 1, the length of the chord covering the reflection line within $\tilde{\Omega}$ is a constant independent of $\sigma_{\tilde{y}}$. This implies that the length of the finite segment covering the reflected images is also constant and independent of $\sigma_{\tilde{y}}$.

3. The coordinate of the point of intersection between the reflection line and

the extension of line 3 (blue point in Figure 4.5) is (x, y) where

$$x = \frac{\sqrt{2}}{2}\left(\frac{-2x_0\rho + 2y_0/\sigma_{\tilde{y}} - 2/\sigma_{\tilde{y}}(1-\rho)}{-2\rho\sqrt{1-\rho}}\right),$$

$$y = \frac{\sqrt{2}}{2}\left(\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\frac{-2x_0\rho + 2y_0/\sigma_{\tilde{y}} - 2/\sigma_{\tilde{y}}(1-\rho)}{-2\rho\sqrt{1-\rho}} + \frac{1}{\sqrt{2}\sigma_{\tilde{y}}\sqrt{1+\rho}}\right).$$

The lengths of the $x-$ and $y-$segments of the chord connecting the initial condition and the point of intersection are

$$\Delta x = \frac{\sqrt{2}}{2}\left\{\frac{1}{\sigma_{\tilde{y}}} \cdot \frac{\sigma_{\tilde{y}}x_0 - y_0/\rho + (1-\rho)/\rho + y_0}{\sqrt{1-\rho}} - \frac{x_0}{\sqrt{1-\rho}}\right\},$$

$$\Delta y = \frac{\sqrt{2}}{2}\left\{\frac{1}{\sigma_{\tilde{y}}} \cdot \frac{(\sigma_{\tilde{y}}x_0 - 2y_0/\rho + (1-\rho)/\rho + 1/\sqrt{2})\sqrt{1-\rho} - y_0\sqrt{1+\rho}}{\sqrt{1-\rho}\sqrt{1+\rho}} - \frac{x_0}{\sqrt{1-\rho}}\right\}.$$

Hence, as $\sigma_{\tilde{y}} \to 0$, both $\Delta x^2 \to \infty$ and $\Delta y^2 \to \infty$, so that the distance between the point of intersection and the initial condition grows. The symmetry of the problem (in terms of reflections $x \to -x$ and $y \to -y$) makes the argument applied to line 3 and the line of reflection applicable to the extension of line 1 as well. Figures 4.7 and 4.6 show the reflected images falling within the interior region with $\sigma_{\tilde{y}} = 0.1$ in both geometries.

Observations ii) and iii) imply that we can always find a sufficiently small $\sigma_{\tilde{y}}$ to cover the line segment containing the positions of images resulting from a finite number of reflections about boundaries 2 and 4. Hence, any subsequent reflections about lines 1 and 3 are guaranteed to place reflected images outside of the computational region Ω.

For any such finite, admissible collection of images, we can make t^* sufficiently small such that the boundary values are numerically enforced on three out of the four boundaries (since with a finite set of reflections the boundaries are analytically enforced at the last boundary of reflection). Therefore, we have a collection of images J whose sum satisfies the initial-boundary problem *and* is non-trivially differentiable with respect to the four boundary parameters for $\tilde{\Omega}$. Moreover, the reflections J cover the required set in (4.4). $\square$

4.4 Matching the Small-time and Galerkin Solutions

So far we have developed an analytic likelihood solution valid for certain small $\tilde{t}_\epsilon$ determined by the geometry of the initial-boundary value problem. The Galerkin likelihood solution from Chapter 3, on the other hand, applies for relatively larger $\tilde{t}$. The results of the simulation studies in Chapter 3 showed the need to resolve the likelihood in the *transient* region $\tilde{t}$ between where the small-time likelihood is applicable and where the Galerkin likelihood is applicable. This is achieved by a two-step process where we first approximate the Galerkin likelihood with a few low-frequency modes via least squares then match the first two derivatives of the small-time likelihood to those of the low-frequency approximation. The resultant interpolation is a matching solution which also resolves the transient region.

The likelihood computed with the Galerkin solution as a function of $\tilde{t}$ is of the form

$$\frac{\partial^4 p_G(\tilde{x}, \tilde{y}, \tilde{t})}{\partial a_x \partial b_x \partial a_y \partial b_y} = \sum_{k=1}^{K} e^{-\lambda_k \tilde{t}} p_k^{(4)}(\tilde{t}),$$

where $p_i^{(4)}(\tilde{t})$ is a fourth-order polynomial. This proceeds from the Galerkin solution being dependent on $\tilde{t}$ only through the exponential term: the eigenfunctions of the solution are by design solely functions of (a_x, b_x, a_y, b_y) and $(\tilde{x}, \tilde{y})$. Expand-

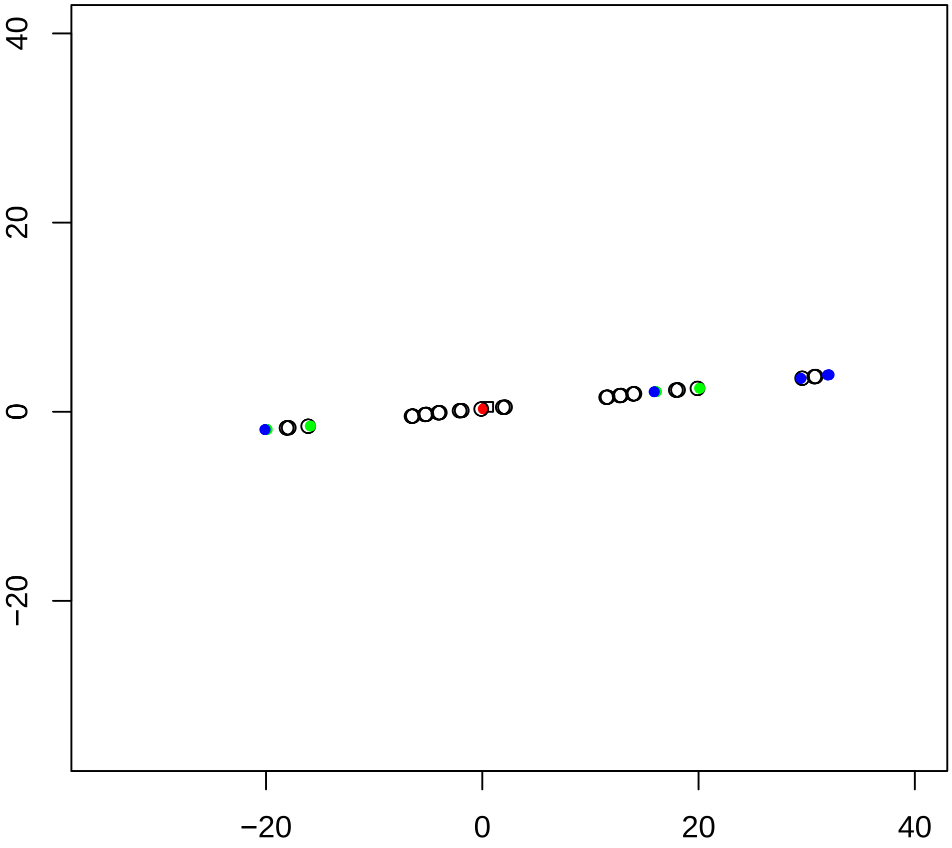

Figure 4.6: Set of images in the original topology when $\sigma_{\tilde{y}} = 0.1$. All images fall outside of $\tilde{\Omega}$ for $\rho = 0.9$.

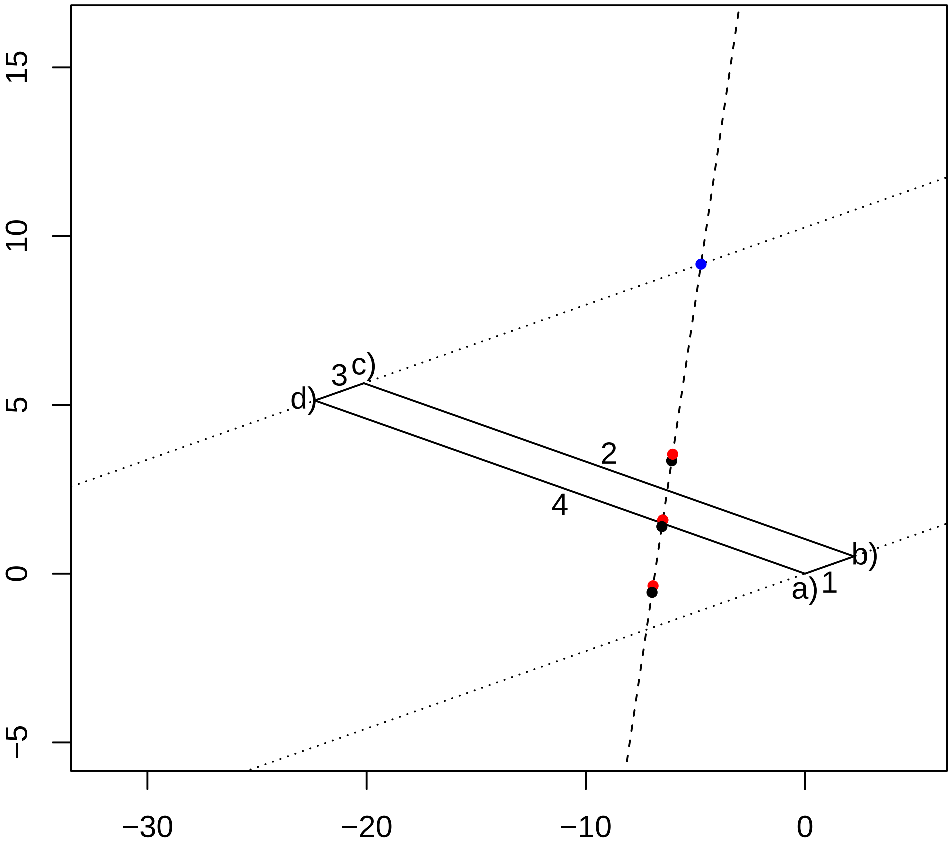

Figure 4.7: Set of images in the transformed topology when $\tilde{\sigma} = 0.1$. All images fall outside of $\tilde{\Omega}$ for $\rho = 0.9$.

ing the polynomials, we see that the generating functions for the likelihood are of the form

$$g_{k,j}(t) := e^{-\lambda_k \tilde{t}}\, t^{\alpha_j}, \qquad \alpha_j = 0, 1, 3, 4. \tag{4.12}$$

The Laplace approximation of $g_{k,j}$ with $\alpha_j > 0$ is useful in elucidating how each of the generating functions contribute to the overall likelihood. Letting

$$\tilde{T} = \log(\tilde{t}), \tag{4.13}$$

to stabilize the tails of the approximant and performing a second-order Taylor expansion of $g(\tilde{T})$ about the maximum of $g_{k,j}$ produces the Gaussian approximation

$$g_{k,j}(\tilde{T}) \approx C(\lambda_k, \alpha_j) \exp\left\{-\frac{1}{2}\alpha_j(\tilde{T} - \log(\alpha_j/\lambda_k))^2\right\}. \tag{4.14}$$

Up to second order, therefore, the Galerkin likelihood (as a function of $\tilde{t}$) is a linear combination of Gaussian kernels $g_{k,j}$ whose bandwidth is controlled by the degree of polynomial multiplier α_j and whose location is controlled by the eigenvalue λ_j and α_j. In particular, the solution exhibits the behavior of resolving smaller $\tilde{t}$ regions with higher-frequency modes. The Laplace approximation of the generating functions is especially appropriate for the transient region of $\tilde{t}$. There, the "non-Gaussian" generating functions $e^{-\lambda_k \tilde{t}}$ tend to unity, while the approximately Gaussian $g_{k,j}$ are dominated by the $\tilde{t}^{\alpha_j}$ terms (left-hand tails of (4.14)).

The first part of the matched solution is constructed from the generating functions $g_{k,j}$ by fitting via least squares to values produced outside the transient region, *i.e.* using big-$\tilde{t}$ likelihood values. We need to use as few points as possible in the fitting exercise, as evaluating the Galerkin likelihood is relatively expensive, while still accurately representing the right-hand tail of the solution in the big-$\tilde{t}$ region. For this reason, we choose a sparse representation of the likelihood using the first two generating functions $g_{1,4}, g_{2,4}$ as proportional to the matched solution

$$\begin{aligned} f_{LS}(\tilde{t}) &= \tilde{t}^4 \left(\omega_1 e^{-\lambda_1 \tilde{t}} + \omega_2 e^{-\lambda_2 \tilde{t}}\right), \\ \log f_{LS}(\tilde{t}) &= 4\log(\tilde{t}) - \lambda_1 \tilde{t} + \log\left(\omega_1 + \omega_2 e^{-(\lambda_2 - \lambda_1)\tilde{t}}\right) \end{aligned}$$

where f_{LS} is asymptotically valid for large $\tilde{t}$. Also, the largest possible degree of $\tilde{t}^\alpha$ used can more easily fit the small-$\tilde{t}$ likelihood which rapidly drops near $\tilde{t} = 0$. The weights ω_1, ω_2 are estimated from two likelihood values outside the transient region. The two points of evaluation are $\tilde{t}_1 = 0.80$ and $\tilde{t}_2 = 1.80$, chosen based on the empirical behavior of the Galerkin solver observed in Figures 3.3 - 3.5. If the Galerkin solver produces an invalid likelihood at $\tilde{t} = 0.80$ or $\tilde{t} = 1.80$, then both trial points are increased by 1 until valid likelihoods are produced. The reason for the large jumps in this case is the relatively high cost of using the Galerkin solver and the guarantee of the existence of a valid likelihood for sufficiently large $\tilde{t}$. Once obtained, the two likelihood values are used to compute the best-fitting ω_1, ω_2 via least squares. This matching scheme is applied to the same points as in Figures 3.3 - 3.5 and are shown in Figures 4.9 - 4.11. This least-squares (LS) solution is used to approximate the location $\tilde{t}_m$ and value of the maximum of the likelihood, as well as compute its first two derivatives there analytically. This data is the right-hand side of the matching solution. We pick this point, because it is close to the maximum under the Galerkin solution and as such it is the inflection point where the dynamics of the solution move from being dominated by the $-\beta/\tilde{t}$ term (as we show below) to those dominated by $-\lambda_1 \tilde{t}$ term. Next, we use the small-time likelihood to pick the functional form of the matching solution that interpolates between the small-time region and the maximum point of the

Galerkin solution, as well as derive the data necessary for the left-hand side of the matching.

Closer observation of (4.11) reveal that each of the summands of the truncated small-time solution share a common polynomial factor:

$$
\begin{aligned}
&\frac{\partial^4 p_\epsilon(\tilde{x},\tilde{y},\tilde{t})}{\partial a_x \partial b_x \partial a_y \partial b_y} \approx \\
&\approx \sum_{j'=1}^{4} \frac{1}{\pi\sqrt{2}} \left(\frac{1}{2\ \tilde{t}\,\sigma_{\tilde{y}}^2(1-\rho^2)}\right)^{4.5} \exp\left(-\frac{1}{2\ \tilde{t}\,\sigma_{\tilde{y}}^2(1-\rho^2)}\mathcal{P}_{j'}\right)\left(\partial_{a_x}\partial_{b_x}\partial_{a_y}\partial_{b_y}\right)\mathcal{P}_{j'} \\
&\qquad = K\left(\frac{1}{\tilde{t}}\right)^{4.5} \sum_{j'=1}^{4} c'_j \exp\left(-\frac{\beta_{j'}}{\tilde{t}}\right), \\
&\qquad K = \frac{1}{\pi\sqrt{2}}\left(\frac{1}{2\ \sigma_{\tilde{y}}^2(1-\rho^2)}\right)^{4.5}, \\
&\qquad \beta_{j'} = \frac{1}{2\ \sigma_{\tilde{y}}^2(1-\rho^2)}\mathcal{P}_{j'}, \\
&\qquad c_{j'} = \left(\partial_{a_x}\partial_{b_x}\partial_{a_y}\partial_{b_y}\right)\mathcal{P}_{j'}
\end{aligned}
$$

The summand with the greatest $\beta_{j'}$ contributes the most to the truncated small-time solution in the $\tilde{t} \leq 1$ region where the matched solution will be applied. Indexing j' such that $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$, the small-time log-likelihood is

$$
\begin{aligned}
\log\left(\frac{\partial^4 p_\epsilon(\tilde{x},\tilde{y},\tilde{t})}{\partial a_x \partial b_x \partial a_y \partial b_y}\right) &\approx \log(K) - 4.5\log(\tilde{t}) + \log(c_1) - \frac{\beta_1}{\tilde{t}} \\
&\quad + \log\left(1 + \sum_{j\neq 1}\frac{c_j}{c_1}\exp\left(-\frac{(\beta_j-\beta_1)}{\tilde{t}}\right)\right) \\
&\approx \log(K) - 4.5\log(\tilde{t}) + \log(c_1) - \frac{\beta_1}{\tilde{t}} + \log\left(1+\epsilon(\tilde{t})\right),
\end{aligned}
\tag{4.15}
$$

where we are assuming that the non-dominant exponential terms are small relative to the β_1 term. When constructing the matched solution, we ignore $\epsilon(\tilde{t})$ since the likelihood function is dominated by the $\log(\tilde{t})$ term away from zero.

The assumed form for the matched solution is the same as (4.15):

$$\log f_{\text{matched}}(\tilde{t}) = \log(\omega(\tilde{t})) - \gamma(\tilde{t})\log(\tilde{t}) - \frac{\beta(\tilde{t})}{\tilde{t}}$$

where the parameters $\omega(\tilde{t}), \gamma(\tilde{t})$, and $\beta(\tilde{t})$ vary with $\tilde{t}$. At $\tilde{t}^*$, the left-hand side of the matching condition, the values for these parameters are defined such that they match the small-time solution

$$\omega(\tilde{t}^*) = K, \qquad \gamma(\tilde{t}^*) = 4.5, \qquad \beta(\tilde{t}^*) = \beta_1.$$

At $\tilde{t}_m$, the right-hand side of the matching condition and the maximum of the LS solution, $\omega(\tilde{t}), \gamma(\tilde{t})$, and $\beta(\tilde{t})$ are chosen to match the value, first, and second derivatives of the logarithmic form of the LS solution. The form of the parameters between $\tilde{t}^*$ and $\tilde{t}_m$ is chosen to be a sigmoid function which rapidly transitions away from $\tilde{t}^*$ and is the same for all three parameters:

$$\begin{aligned}
\omega(\tilde{t}) &= \omega(\tilde{t}^*)e^{-k(\tilde{t}-\tilde{t}^*)} + \omega(\tilde{t}_m)\left(1 - e^{-k(\tilde{t}-\tilde{t}^*)}\right), \\
\gamma(\tilde{t}) &= \gamma(\tilde{t}^*)e^{-k(\tilde{t}-\tilde{t}^*)} + \gamma(\tilde{t}_m)\left(1 - e^{-k(\tilde{t}-\tilde{t}^*)}\right), \\
\beta(\tilde{t}) &= \beta(\tilde{t}^*)e^{-k(\tilde{t}-\tilde{t}^*)} + \beta(\tilde{t}_m)\left(1 - e^{-k(\tilde{t}-\tilde{t}^*)}\right).
\end{aligned} \tag{4.16}$$

Given the order magnitudes of $\tilde{t}^*$ and $\tilde{t}_m$ commonly encountered, $k \approx 100$ is sufficiently large such that the higher-order derivatives of $1 - e^{-k(\tilde{t}-\tilde{t}^*)}$ are approximately zero at $\tilde{t}_m$ and the matching on the right-hand side is still satisfied.

4.5 MLEs, revisited

To complete the analysis, we consider again the simulation scenario in Chapter 3 and the experimental setup where the Galerkin solution alone failed to outper-

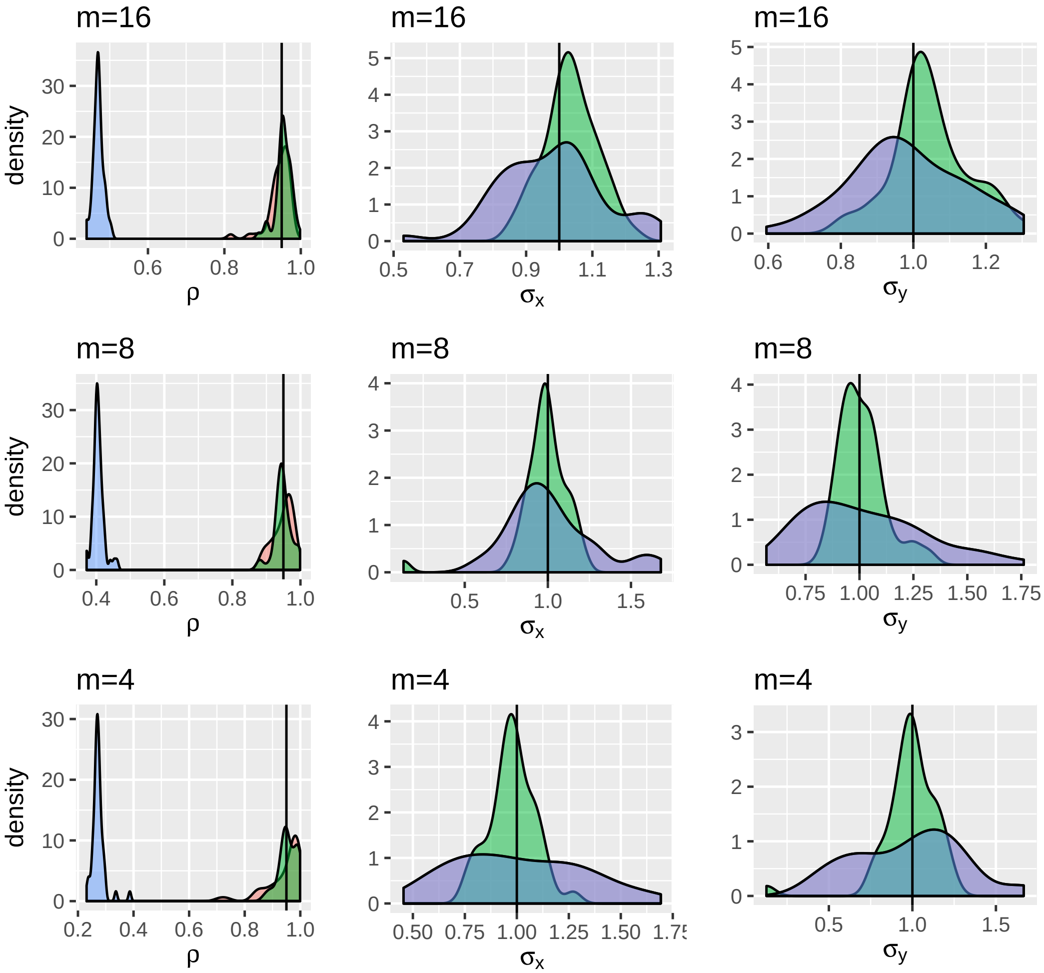

Figure 4.8: Data generated with $\rho = 0.95$ and is the same as in Figure 3.6. Kernel-density approximations of the repeated-sampling densities of the MLEs are shown. Samples are obtained from the Galerkin likelihood with the matching solution (green) and the classical likelihood (red) and the Rogers estimator (blue). The data-generating parameters are denoted with the vertical solid line. Compared to Figure 3.6, the repeated sampling distributions of the parameter estimates based on the Galerkin/matching solution are tighter.

	$\rho = 0.95$		
	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.124	0.429	0.318
$\hat{\sigma}_y$	0.310	0.147	0.365
$\hat{\rho}$	0.250	0.753	0.699

Table 4.1: Ratios of Galerkin to Gaussian MSEs for the three simulation cases.

form the Gaussian (non-boundary) likelihood. In particular, $\rho = 0.95$ with $m = 8$, and 16 showed a greater MSE than the MLE estimates using a purely Gaussian likelihood (see Table 3.1). Using the same simulated data sets, the repeated sampling distribution approximations where the matching solution is used in addition to the Galerkin likelihood are shown in Figure 4.8. The ratios of the MSEs between the matched and Gaussian likelihoods are shown in Table 4.1. We can see that the MSEs are resoundingly better, and that using the matching technique developed here improves the statistical power of using OCHL data.

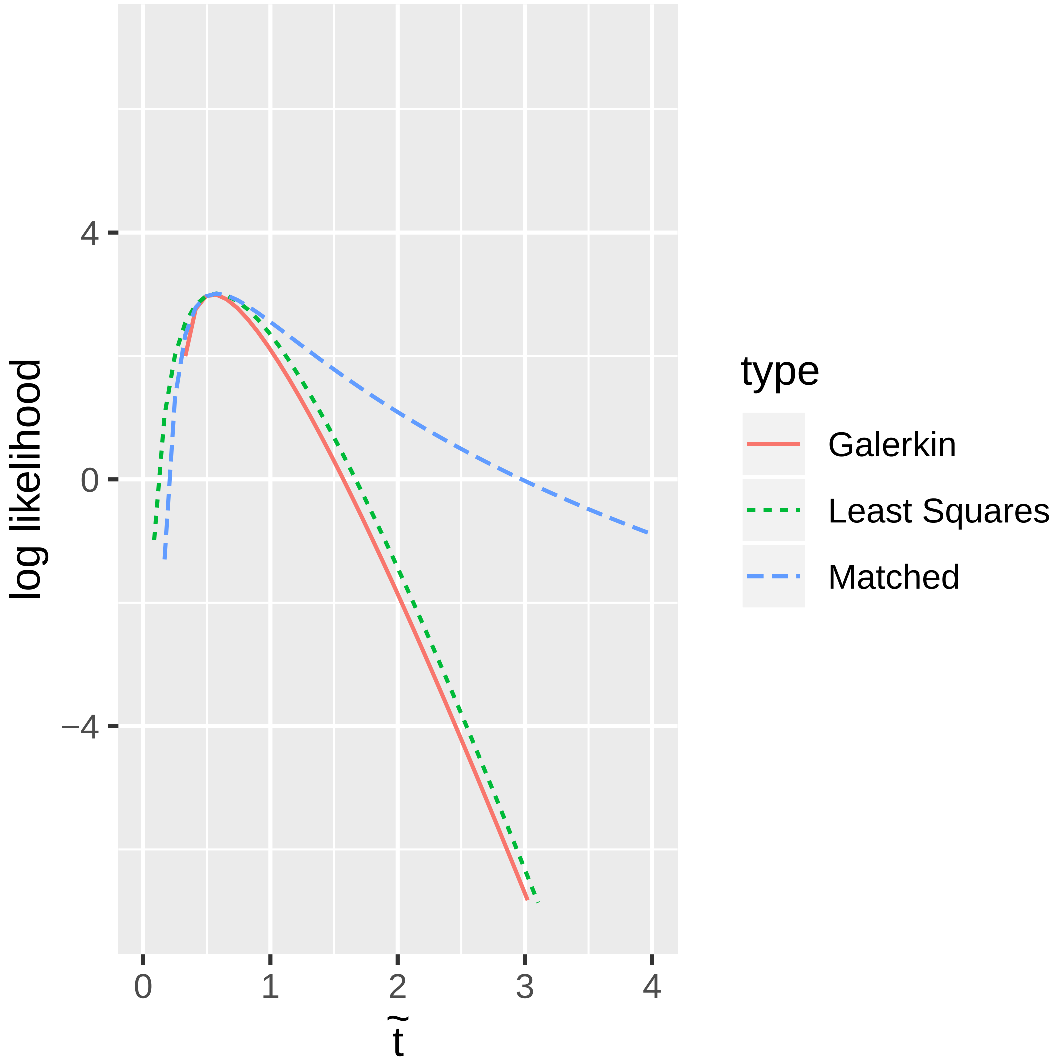

Figure 4.9: Galerkin and matched log-likelihoods for the point in Figure 3.3, computed with different ($\tilde{\rho} = 0.60, \tilde{\sigma} = 0.08$). Also shown are the least-squares solution and the matched solution. We see that the location of the maximum of the least-squares solution is very close to the maximum of the Galerkin solution. The matched solution preserves the asymptotic behavior of the small-time likelihood but also matches the Galerkin likelihood at the peak.

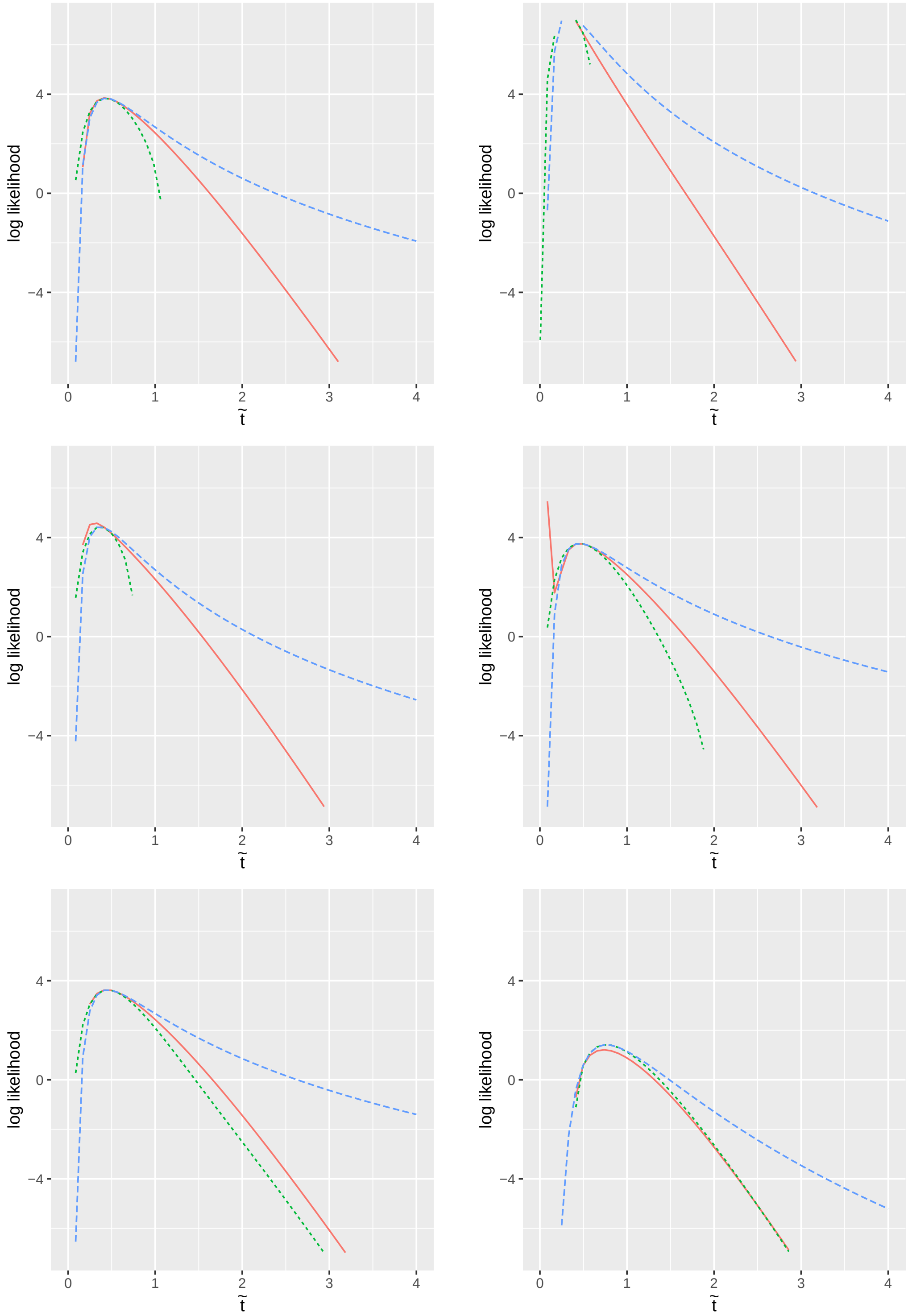

Figure 4.10: Data points 2 through 12 used in Figure 3.4 with the Galerkin solution and the matched asymptotic solution in the transient region. The parameters for the Galerkin solution are $(\tilde{\rho} = 0.60, \tilde{\sigma} = 0.08, l = 1)$

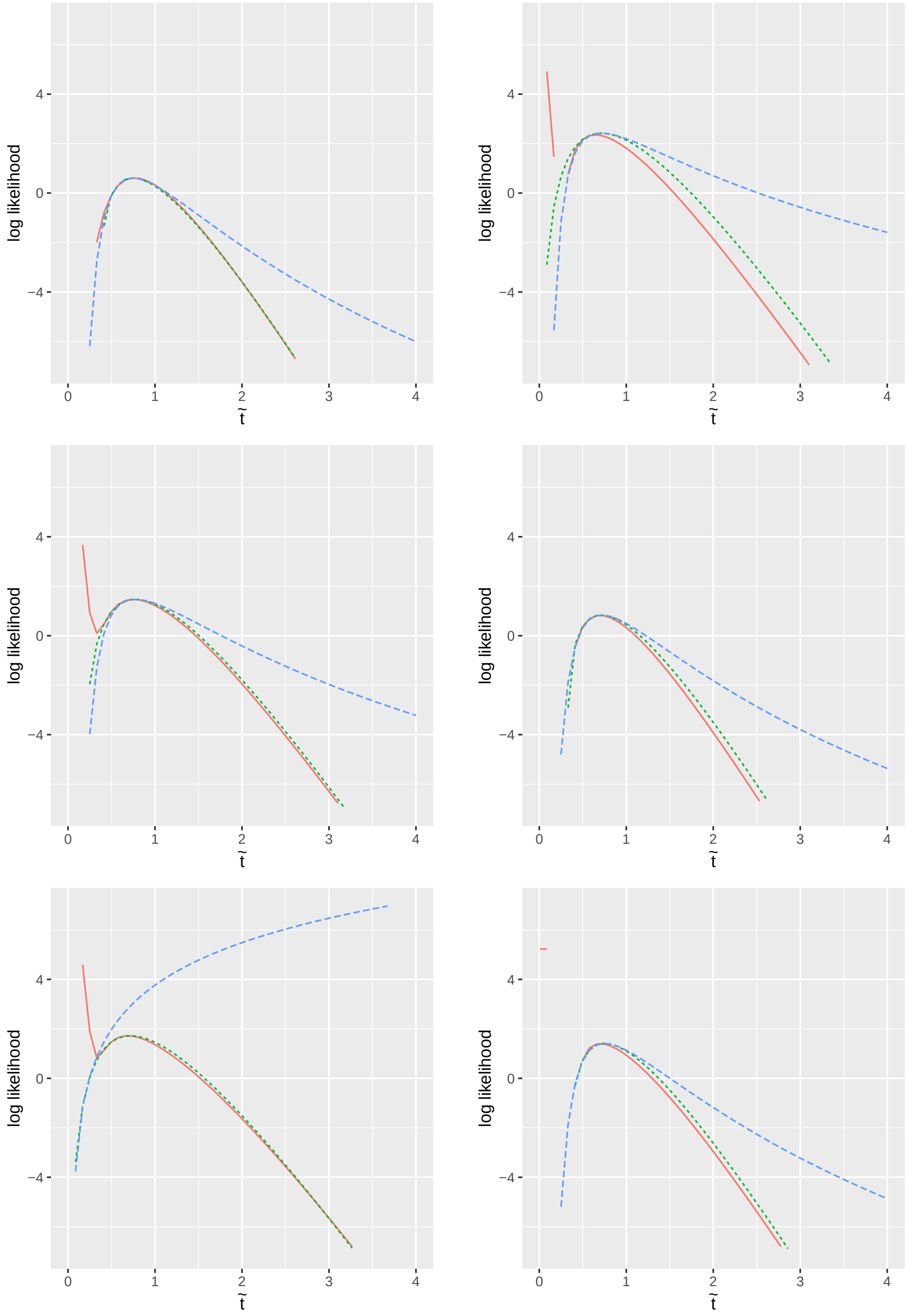

Figure 4.11: Data points 8 through 13 used in Figure 3.5 with the Galerkin solution and the matched asymptotic solution in the transient region. The parameters for the Galerkin solution are $(\tilde{\rho} = 0.60, \tilde{\sigma} = 0.08, l = 1)$

Chapter 5

Conclusion

We have presented novel approaches for using two types of financial data for statistical inference and prediction. We have demonstrated the effectiveness of a Bayesian filtering approach in estimating financial market volatility in noisy, high-frequency data. This method outperforms the currently popular averaging approaches found in the realized volatility literature in estimating integrated volatility on a daily timescale. As part of our approach, we have elicited priors invariant with respect to sampling frequency, and have formulated the used discrete-time state-space models to be coherent over sampling frequencies as well. Future work on this project includes extending the problem to multiple assets. Given the nature of volatility shocks in markets, namely their occurrence over a single or a few days, the high-frequency resolution of the volatility paths and potentially fast-changing correlations among assets are of interest to both practitioners and academics in the field.

We have also developed a semi-analytic computational framework for volatility estimation for bivariate assets with OCHL data, a completely novel contribution. As part of our approach, we have developed a non-separable variational-form basis expansion (Galerkin solution) to the governing differential equation. Our numerical Galerkin solver takes directly into account the correlation structure inherent in the bivariate OCHL problem and is therefore more efficient when compared to other proposed numerical methods. However, in the context statistical inference, often the likelihood in question needs to be evaluated in parameter regions where near-infinite computational memory and time needs to be employed in order to produce a valid numerical solution. To deal with this difficulty, we in turn developed an analytic solution asymptotically valid in this (small-time, under our proposed representation of the canonical problem) parameter region. This was achieved via the method of images and a symmetry argument. Next, we proposed a matching solution which interpolates between the regions where the Galerkin and small-time solutions are applicable. Finally, we performed a series of MLE exercises that showed the validity of our method and also demonstrated the increase in statistical power in using OCHL data in comparison to both the classical Gaussian MLE estimator that disregards boundary conditions as well as the other existing bivariate OCHL estimator.

A natural extension of our OCHL solution is its application in more dynamical settings. In particular, when combined with a Gaussian process emulator, it is possible to implement the OCHL likelihood solution in the estimation of stochastic volatility dynamical models with the use of particle filters.

Appendix A

Details of the Markov chain Monte Carlo algorithm

We rewrite the full discrete-time model

$$\begin{aligned}
Y_j &= \log(S_j) + \zeta_j, \\
\log(S_j) &= \log(S_{j-1}) + \mu(\Delta) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\,\epsilon_j + J_j(\Delta), \\
\log(\sigma_{j+1,1}) &= \alpha(\Delta) + \theta_1(\Delta)\,\{\log(\sigma_{j,1}) - \alpha(\Delta)\} + \tau_1(\Delta)\,\epsilon_{j,1}, \\
\log(\sigma_{j+1,2}) &= \alpha(\Delta) + \theta_2(\Delta)\,\{\log(\sigma_{j,2}) - \alpha(\Delta)\} + \tau_2(\Delta)\,\epsilon_{j,2},
\end{aligned} \tag{A.1}$$

This model is nonlinear in terms of the volatility due to the formulation of its evolution on the log-scale. To re-parameterize the model to be linear in terms of volatility and thereby use the Kalman Filter and Sampler, we take equation (A.1) and transform it so that it is linear in terms of $\log(\sigma_j)$,

$$
\begin{aligned}
&\log(S_j) = \log(S_{j-1}) + \mu(\Delta) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\,\epsilon_j + J_j(\Delta) \\
\leftrightarrow \quad & \underbrace{\log\left[\left|\log(S_j/S_{j-1}) - \mu(\Delta) - J_j(\Delta)\right|\right]}_{y_j^*} = \frac{1}{2}\underbrace{\log(\sigma_{j,1})}_{h_{j,1}} + \frac{1}{2}\underbrace{\log(\sigma_{j,2})}_{h_{j,2}} + \underbrace{\log(\epsilon_{j,1}^2)/2}_{\epsilon_j^*}.
\end{aligned}
$$

Having defined y^*, $h_{j,1}$, $h_{j,2}$ and ϵ_j^*, the model becomes linear in the terms involving the volatility:

$$
\begin{aligned}
Y_j &= \log(S_j) + \zeta_j, & \text{(A.2)} \\
y_j^* &= \frac{1}{2}h_{j,1} + \frac{1}{2}h_{j,2} + \epsilon_j^*, & \text{(A.3)} \\
h_{j+1,1} &= \alpha(\Delta) + \theta_1(\Delta)\left\{h_{j,1} - \alpha(\Delta)\right\} + \tau_1(\Delta)\epsilon_{j,1}, & \text{(A.4)} \\
h_{j+1,2} &= \alpha(\Delta) + \theta_2(\Delta)\left\{h_{j,2} - \alpha(\Delta)\right\} + \tau_2(\Delta)\epsilon_{j,2}, & \text{(A.5)}
\end{aligned}
$$

We approximate ϵ_j^* as a mixture of Normals

$$
\epsilon_j^* = \log(\epsilon_j^2)/2 \sim \sum_{l=1}^{10} p_l N\left(\frac{m_l}{2}, \frac{v_l^2}{4}\right).
$$

We can introduce the mixture indicators $\gamma_1, \ldots, \gamma_{n(\Delta)}$ such that

$$
\log(\epsilon_j^2)/2 \mid \gamma_j \sim N\left(\frac{m_{\gamma_j}}{2}, \frac{v_{\gamma_j}^2}{4}\right), \qquad\qquad \Pr(\gamma_k = l) = p_l.
$$

Hence, conditionally on the sampled prices, the indicators $\gamma_1, \ldots, \gamma_{n(\Delta)}$, jumps $J_1(\Delta)$, $\ldots$, $J_{n(\Delta)}(\Delta)$, and the parameters $\mu(\Delta)$, $\alpha(\Delta)$, $\theta_k(\Delta)$, $\tau_k(\Delta)$ and ρ, we have again a linear state-space model with Gaussian innovations. However, due to correlation of the innovations of the price process ϵ_j and the fast volatility process $\epsilon_{j,2}$, we need a joint distribution for the transformed and approximated ϵ_j^* and $\epsilon_{j,2}$. To this end, we directly follow the approach in Omori et al. [2007], beginning with the known expression

$$\begin{aligned} p(\epsilon_{j,2}, \epsilon_j^*|\gamma_j) &= p(\epsilon_{j,2}|\epsilon_j^*, \gamma_j)p(\epsilon_j^*|\gamma_j) \\ &= p(\epsilon_{j,2}|\underbrace{d_j \exp(\epsilon_j^*)}_{\epsilon_j}, \gamma_j)p(\epsilon_j^*|\gamma_j) \\ &= N\left(\epsilon_{j,2} \,\middle|\, \rho d_j \exp(\epsilon_j^*), (1-\rho^2)\right) N\left(\epsilon_j^* \,\middle|\, \frac{m_{\gamma_j}}{2}, \frac{v_{\gamma_j}^2}{4}\right), \end{aligned}$$

where d_j is the sign of ϵ_j. The nonlinear term $\exp(\epsilon_j^*)$ is approximated by a linear function, where the constants $(a_{\gamma_j}, b_{\gamma_j})$ are chosen to minimize the expected squared difference between $\exp(\epsilon_j^*)$ and its approximation, as done in Omori et al. [2007]

$$\exp(\epsilon_j^*)|\gamma_j \approx \exp(m_{\gamma_j}/2)(a_{\gamma_j} + b_{\gamma_j}(2\epsilon_j^* - m_{\gamma_j})).$$

If $z_j^*, z_j \overset{iid}{\sim} N(0,1)$, the joint distribution for the conditional distribution of the pair $(\epsilon_j^*, \epsilon_{j,2}|\gamma_j)$ can be written as

$$\left\{ \left(\begin{array}{c} \epsilon_j^* \\ \epsilon_{j,2} \end{array} \right) \middle| \gamma_j \right\} = \left(\begin{array}{c} m_{\gamma_j}/2 \\ d_j \rho \exp(m_{\gamma_j}/2) a_{\gamma_j} \end{array} \right) + \left(\begin{array}{cc} v_{\gamma_j}/2 & 0 \\ d_j \rho b_{\gamma_j} v_{\gamma_j} \exp(m_{\gamma_j}/2) & \sqrt{1-\rho^2} \end{array} \right) \left(\begin{array}{c} z_j^* \\ z_j \end{array} \right). \tag{A.6}$$

Rearranging equation (A.3) to express z_j^* in terms of $y_j^*, h_{j,1}, h_{j,2}, m_{\gamma_j}$, and v_{γ_j}, then substituting into equation (A.5) allows us to finally write down the model in the convenient linear state-space form, where the innovations in the state-evolution equations are independent:

$$\begin{aligned} Y_j &= \log(S_j) + \zeta_j, \\ y_j^* &= \frac{1}{2}h_{j,1} + \frac{1}{2}h_{j,2} + \frac{m_{\gamma_j}}{2} + \frac{v_{\gamma_j}}{2}\ z_j^*, \\ h_{j+1,1} &= \theta_1(\Delta)h_{j,1} + \alpha(\Delta)(1-\theta_1(\Delta)) + \tau_1(\Delta)\epsilon_{j,1}, \\ h_{j+1,2} &= \theta_{j,1}(\Delta)h_{j,1} + \theta_{j,2}(\Delta)h_{j,2} + \alpha_j(\Delta) + \tau_2(\Delta)\sqrt{1-\rho^2}z_j, \\ \Rightarrow \begin{pmatrix} h_{j,1} \\ h_{j,2} \end{pmatrix} &= G_j(\Delta)\begin{pmatrix} h_{j,1} \\ h_{j,2} \end{pmatrix} + \begin{pmatrix} \alpha(\Delta)(1-\theta_1(\Delta)) \\ \alpha_j(\Delta) \end{pmatrix} + C(\Delta)\begin{pmatrix} \epsilon_{j,1} \\ z_j \end{pmatrix} \end{aligned} \tag{A.7}$$

with

$$\begin{aligned} G_j(\Delta) &= \begin{pmatrix} \theta_1(\Delta) & 0 \\ \theta_{j,1}(\Delta) & \theta_{j,2}(\Delta) \end{pmatrix} \\ \theta_{j,1}(\Delta) &= -\tau_2(\Delta)d_j\rho b_{\gamma_j}\exp(m_{\gamma_j}/2) \\ \theta_{j,2}(\Delta) &= \theta_2(\Delta) - \tau_2(\Delta)d_j\rho b_{\gamma_j}\exp(m_{\gamma_j}/2) \\ \alpha_j(\Delta) &= \alpha(\Delta)(1-\theta_2(\Delta)) + \\ &\qquad \tau_2(\Delta)\left(d_j\rho\exp(m_{\gamma_j}/2)a_{\gamma_j} + d_j\rho b_{\gamma_j}v_{\gamma_j}\exp(m_{\gamma_j}/2)\left\{\frac{y_j^* - m_{\gamma_j}/2}{v_{\gamma_j}/2}\right\}\right) \\ C(\Delta) &= \begin{pmatrix} \tau_1(\Delta) & 0 \\ 0 & \tau_2(\Delta)\sqrt{1-\rho^2} \end{pmatrix} \end{aligned}$$

The full likelihood for the model can be written as

$$
\begin{aligned}
p(Y_1,\ldots,Y_{n(\Delta)}|h_{1,\ldots,n(\Delta),n(\Delta)+1},\log(S_{0,\ldots,n(\Delta)}),\gamma_{1,\ldots,n(\Delta)},J_{1,\ldots,n(\Delta)}(\Delta),\boldsymbol{\psi}) \propto \\
\prod_{j=1}^{n(\Delta)} \xi^{-1/2}\exp\left\{-\frac{1}{2\xi}(Y_j-\log(S_j))^2\right\} \\
\times \prod_{j=1}^{n(\Delta)} (v_{\gamma_j}/2)^{-1}\exp\left\{-\frac{1}{2v_{\gamma_j}^2/4}\left(y_j^*-h_{j,1}/2-h_{j,2}/2-m_{\gamma_j}/2\right)^2\right\} \\
\times \prod_{j=1}^{n(\Delta)} (\tau_1(\Delta))^{-1}\exp\left\{-\frac{1}{2\tau_1(\Delta)^2}\left(h_{j+1,1}-\theta_1(\Delta)h_{j,1}-\alpha(\Delta)\right)^2\right\} \\
\times \prod_{j=1}^{n(\Delta)} \left(\tau_2(\Delta)\sqrt{1-\rho^2}\right)^{-1}\exp\left\{-\frac{1}{2\tau_2(\Delta)^2(1-\rho^2)}\left(h_{j+1,2}-\theta_{j,2}(\Delta)h_{j,2}-\alpha_{j,2}(\Delta)\right)^2\right\} \\
\times p(\log(S_0))p(\log(\sigma_1)),
\end{aligned}
$$

where $\boldsymbol{\psi}=(\xi,\mu(\Delta),\rho,\tau_1(\Delta),\tau_2(\Delta),\theta_1(\Delta),\theta_2(\Delta),\alpha(\Delta))$.

For our MCMC algorithm, we implement a Gibbs sampler where we simulate posterior draws from the full conditional posteriors for each set of parameters in the steps below. We will use the following abbreviations for ease of notation:

$$
\begin{aligned}
\boldsymbol{\gamma} &:= (\gamma_1,\ldots,\gamma_{n(\Delta)}), & \log(\boldsymbol{S}) &:= (\log(S_1),\ldots,\log(S_{n(\Delta)})), \\
\boldsymbol{\sigma} &:= \left((\sigma_{1,1},\sigma_{1,2}),\ldots,(\sigma_{n(\Delta)+1,1},\sigma_{n(\Delta)+1,2})\right), & \boldsymbol{Y} &:= (Y_1,\ldots,Y_{n(\Delta)}), \\
\boldsymbol{J} &:= (J_1(\Delta),\ldots,J_{n(\Delta)}), & \boldsymbol{\Omega} &:= (\mu(\Delta),\rho,\xi^2), \\
\boldsymbol{\theta} &:= (\alpha(\Delta),\theta_1(\Delta),\theta_2(\Delta),\tau_1(\Delta),\tau_2(\Delta)), & \boldsymbol{\Lambda} &:= (\lambda,\mu_J,\sigma_J^2).
\end{aligned}
$$

1. **Sample $\boldsymbol{p}(\log(\boldsymbol{S}),\boldsymbol{\Omega}|\boldsymbol{\Lambda},\boldsymbol{\theta},\boldsymbol{J},\boldsymbol{\sigma},\boldsymbol{Y})$.**

   (a) **Sample $\boldsymbol{p}(\boldsymbol{\Omega}|\boldsymbol{\Lambda},\boldsymbol{\theta},\boldsymbol{J},\boldsymbol{\sigma},\boldsymbol{Y})$.** Conditional on the volatilities/jumps $\boldsymbol{\sigma}$ and $\boldsymbol{J}$ and volatility/jump parameters $\boldsymbol{\theta},\boldsymbol{\Lambda}$, the discrete-time version of model in (A.1) is comprised of the linear system

$$\begin{aligned}
Y_j &= \log(S_j) + \zeta_j, \\
\log(S_j) &= \mu(\Delta) + \log(S_{j-1}) + J_j(\Delta), \\
&\quad + \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2} + \sqrt{(1-\rho^2)\sigma_{j,1}\sigma_{j,2}} \cdot \epsilon_j, \\
\epsilon_{j,2} &= \frac{\log(\sigma_{j+1,1}) - \alpha(\Delta) - \theta_2(\Delta)\left\{\log(\sigma_{j,1}) - \alpha(\Delta)\right\}}{\tau_2(\Delta)}, \qquad (A.8) \\
\epsilon_j &\sim N(0,1), \\
\zeta_j &\sim N(0,\xi^2), \\
p(\log(S_0)) &= N\left(\log(S_0) \,\middle|\, \eta, \kappa^2\right).
\end{aligned}$$

The observational likelihood model in (A.8) allows us to integrate out out $\log(\boldsymbol{S})$ using a sequential application of Bayes' Theorem forward in time, i.e. the Forward Filter. Note that starting with (A.1) is equivalent to also integrating out $\boldsymbol{\gamma}$ from the system as well. Hence we arrive at the conditional likelihood for observations $p(\boldsymbol{Y}|\boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{J}, \boldsymbol{\sigma}, \boldsymbol{\theta})$ with which we can sample from the posterior:

$$p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{J}, \boldsymbol{\sigma}, \boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\theta}, \boldsymbol{J}, \boldsymbol{\sigma}, \boldsymbol{Y})p(\boldsymbol{\Omega}).$$

This is accomplished with a random-walk Metropolis-Hastings step whose proposal covariance matrix is tuned to the data. In particular, sampling as well as accepting/rejecting is done on the $\mathbb{R}^3$ scale via the transformation $\boldsymbol{\Omega} \to \tilde{\boldsymbol{\Omega}} := \left(\mu(\Delta), \text{logit}\left(\frac{\rho+1}{2}\right), \log(\xi^2)\right)$.

(b) **Sample $\boldsymbol{p(}\log(\boldsymbol{S})|\boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\theta}, \boldsymbol{J}, \boldsymbol{\sigma}, \boldsymbol{Y}\boldsymbol{)}$.** The latent log-prices are sampled using the Forward Filter (as in step 1(a) above) and Backward Sampler using the system in (A.8). Because each y_j^* is dependent on $\log(S_j)$ we re-define y_j^* after this sample.

2. **Sample the jumps and jump parameters: $\boldsymbol{p(J, \Lambda | \Omega, \theta, \log(S), \sigma, Y)}$**

   (a) We can integrate out $\boldsymbol{J}$ from the likelihood model in (A.1), (A.8) by considering the likelihood of occurrence of jumps

   $$p(N(\Delta) = 0) = e^{-\lambda\Delta}, \qquad p(N(\Delta) > 0) = 1 - e^{-\lambda\Delta}.$$

   We assume that Δ is sufficiently small such that only a single jump occurs within the j^{th} period, such that the full conditional distribution of $J_j(\Delta)$ is therefore

   $$p(J_j(\Delta)|\boldsymbol{\Lambda}) = e^{-\lambda\Delta} \cdot \mathbb{1}_{(N(\Delta)=0)} + (1 - e^{-\lambda\Delta}) \cdot N(J_j(\Delta)|\mu_j, \sigma_J^2) \cdot \mathbb{1}_{(N(\Delta)>0)}.$$

   Integrating out the realized $J_j(\Delta)$ from (A.8) produces the likelihood model for each $\log(S_j)$:

   $$\begin{aligned} p(\log(\boldsymbol{S})|\boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\sigma}) = &\prod_{j=1}^{n(\Delta)} p(\log(S_j)|\log(S_{j-1}), \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\sigma}) \\ &\times p(\log(S_0)), \end{aligned} \tag{A.9}$$

$$
\begin{aligned}
&p(\log(S_j)|\log(S_{j-1}), \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\sigma}) = \\
&\qquad e^{-\lambda\Delta}\cdot \\
&N\left(\log(S_j)|\mu(\Delta) + \log(S_{j-1}) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2}, \sigma_{j,1}\sigma_{j,2}(1-\rho^2)\right) \\
&\qquad + (1-e^{-\lambda\Delta})\cdot \\
&N\left(\log(S_j)|\mu(\Delta) + \log(S_{j-1}) + \mu_J + \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2}, \sigma_J^2 + \sigma_{j,1}\sigma_{j,2}(1-\rho^2)\right).
\end{aligned}
\tag{A.10}
$$

Given that there is no dependence on $\boldsymbol{\Lambda}$ in the observable model, the posterior distribution for $\boldsymbol{\Lambda}$ is computable with (A.9) - (A.10) via the relation

$$
p(\boldsymbol{\Lambda}|\boldsymbol{\Omega}, \boldsymbol{\theta}, \log(\boldsymbol{S}), \boldsymbol{\sigma}, \boldsymbol{Y}) \propto p(\log(\boldsymbol{S})|\boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\sigma})p(\boldsymbol{\Lambda}.)
$$

As in step 1a, we use a random-walk Metropolis-Hastings step to sample $\boldsymbol{\Lambda}$ from the posterior.

(b) **Sample $\boldsymbol{p(J|\Lambda, \Omega, \theta,} \log\boldsymbol{(S), \sigma, Y)}$.** The posterior probability of the indicator function on each jump is

$$
\begin{aligned}
\pi_0 &:= p(N_j(\Delta) = 0|\boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \log(\boldsymbol{S}), \boldsymbol{\sigma}, \boldsymbol{Y}) \\
&\propto N\left(\log(S_j)|\mu(\Delta) + \log(S_{j-1}) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2}, \sigma_{j,1}\sigma_{j,2}(1-\rho^2)\right) \\
&\quad \cdot \exp(-\lambda\Delta) \\
\pi_1 &:= p(N_j(\Delta) > 0|\boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \log(\boldsymbol{S}), \boldsymbol{\sigma}, \boldsymbol{Y}) \\
&\propto N\left(\log(S_j)|\mu(\Delta) + \log(S_{j-1}) + \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2} + \mu_J, \sigma_{j,1}\sigma_{j,2}(1-\rho^2) + \sigma_J^2\right) \\
&\qquad \cdot (1 - \exp(-\lambda\Delta))
\end{aligned}
$$

Given a non-zero jump at time j, the posterior distribution of its size is proportional to a Normal distribution

$$
\begin{aligned}
& p(J_j(\Delta)|\boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \log(\boldsymbol{S}), \boldsymbol{\sigma}, \boldsymbol{Y}) \propto \\
& N\left(J_j(\Delta)|\log(S_j) - \mu(\Delta) - \log(S_{j-1}) - \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2}, \sigma_{j,1}\sigma_{j,2}(1-\rho^2)\right) \\
& \qquad \cdot N(J_j(\Delta)|\mu_J, \sigma_J^2) \\
& \propto N\left(J_j(\Delta)\middle|\left(\frac{\log(S_j) - \mu(\Delta) - \log(S_{j-1}) - \sqrt{\sigma_{j,1}\sigma_{j,2}}\rho\epsilon_{j,2}}{\sigma_{j,1}\sigma_{j,2}(1-\rho^2)} + \frac{\mu_J}{\sigma_J^2}\right)\cdot\right. \\
& \qquad \left(\frac{1}{(1-\rho^2)\sigma_{j,1}\sigma_{j,2}} + \frac{1}{\sigma_J^2}\right)^{-1}, \\
& \qquad \left.\left(\frac{1}{(1-\rho^2)\sigma_{j,1}\sigma_{j,2}} + \frac{1}{\sigma_J^2}\right)^{-1}\right)
\end{aligned}
$$

3. **Sample $\boldsymbol{p(\gamma|\sigma, \theta, \Lambda, \Omega, J, \log(S), Y)}$.** Before sampling $\boldsymbol{\gamma}$, all y_j^* must be re-computed. Following that, since each γ_j can take on a finite number of values, for each j we sample the discrete posterior where

$$
\begin{aligned}
p(\gamma_j = l|-) \propto\ & p(\gamma = l)(v_l/2)^{-1}\exp\left\{-\frac{1}{2v_l^2/4}(y_j^* - h_{j,1}/2 - h_{j,2}/2 - m_l/2)^2\right\} \\
& \times \exp\left\{-\frac{1}{2\tau_2(\Delta)^2(1-\rho^2)}\left(h_{j+1,2} - \theta_{j,1}(\Delta)h_{j,1} - \theta_{j,2}(\Delta)h_{j,2} - \alpha_{j,2}(\Delta)\right)^2\right\}
\end{aligned}
$$

4. **Sample the volatility parameters $\boldsymbol{p(\sigma, \theta|\gamma, \Lambda, \Omega, \log(S), Y)}$**

   (a) **Sample $\boldsymbol{p(\theta|\gamma, \Lambda, \Omega, \log(S), Y)}$.** Conditional on all other parameters, the portion of the state-space model where $h_{j,k}$ appear is comprised of the linear system

$$
\begin{aligned}
y_j^* &= \frac{1}{2}h_{j,1} + \frac{1}{2}h_{j,2} + \frac{m_{\gamma_j}}{2} + \frac{v_{\gamma_j}}{2}\ z_j^*, \\
h_{j+1,1} &= \theta_1(\Delta)h_{j,1} + \alpha(\Delta) + \tau_1(\Delta)\epsilon_{j,1}, \qquad \text{(A.11)} \\
h_{j+1,2} &= \theta_{j,2}(\Delta)h_{j,2} + \alpha_{j,2}(\Delta) + \tau_2(\Delta)\sqrt{1-\rho^2}z_j,
\end{aligned}
$$

with $\alpha_{j,2}(\Delta)$ and $\theta_{j,2}(\Delta)$ defined below (A.7). Thus we can integrate out $h_{j,k}$ using the Forward Filter and obtain the likelihood

$$p(y_1^*, \dots, y_{n(\Delta)}^* | \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Lambda}, \boldsymbol{\Omega})$$

and sample from the posterior for $\boldsymbol{\theta}$ using a Metropolis-Hastings step. As before, we transform $\boldsymbol{\theta} \to \tilde{\boldsymbol{\theta}}$ such that $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^5$ and perform sampling on this scale.

(b) **Sample $\boldsymbol{p(\sigma|\theta, \gamma, \Lambda, \Omega, \log(S), Y)}$.** Conditional on $\boldsymbol{\theta}$ and all other parameters, we use the Forward Filter Backward Sampler on (A.11) to sample $\boldsymbol{\sigma}$.

Bibliography

Vyacheslav M Abramov and Fima C Klebaner. Estimation and prediction of a non-constant volatility. *Asia-Pacific Financial Markets*, 14(1):1–23, 2007.

Deepak K Agarwal and Alan E Gelfand. Slice sampling for simulation based fitting of spatial data models. *Statistics and Computing*, 15(1):61–69, 2005.

Yacine Ait-Sahalia, Per A Mykland, and Lan Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175, 2011.

S. Alizadeh, M. Brandt, and F. Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57:1047–1092, 2002a.

Sassan Alizadeh, Michael W Brandt, and Francis X Diebold. Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091, 2002b.

Torben G Andersen and Tim Bollerslev. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The journal of Finance*, 52(3):975–1005, 1997a.

Torben G Andersen and Tim Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2):115–158, 1997b.

Torben G Andersen and Tim Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905, 1998.

Torben G Andersen, Tim Bollerslev, and Steve Lange. Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5):457–477, 1999.

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Heiko Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76, 2001a.

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55, 2001b.

Manabu Asai, Michael McAleer, and Jun Yu. Multivariate stochastic volatility: a review. *Econometric Reviews*, 25(2-3):145–175, 2006.

Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

Xuezheng Bai. *Beyond Merton's utopia: effects of non-normality and dependence on the precision of variance estimaters using high-frequency financial data*. PhD thesis, University of Chicago, Graduate School of Business., 2000.

Richard T Baillie and Robert J Myers. Bivariate garch estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, 6(2):109–124, 1991.

Ole E Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.

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

Ole E Barndorff-Nielsen and Neil Shephard. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477, 2002a.

Ole E Barndorff-Nielsen and Neil Shephard. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477, 2002b.

Ole E Barndorff-Nielsen and Neil Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37, 2004.

Ole E Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.

David S Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies*, 9(1):69–107, 1996.

Luc Bauwens and Michel Lubrano. Bayesian inference on garch models using the gibbs sampler. *The Econometrics Journal*, 1(1):23–46, 1998.

Luc Bauwens, Sébastien Laurent, and Jeroen VK Rombouts. Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109, 2006.

Fischer Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1):167–179, 1976.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics*, pages 498–505, 1990.

Tim Bollerslev and Jonathan H Wright. High-frequency data, frequency domain inference, and volatility forecasting. *Review of Economics and Statistics*, 83(4): 596–602, 2001.

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

Tim Bollerslev and Hao Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1): 33–65, 2002.

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.

Michael W Brandt and Francis X Diebold. A no-arbitrage approach to range-based estimation of return covariances and correlations. Technical report, National Bureau of Economic Research, 2003.

Michael W Brandt and Pedro Santa-Clara. Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61(5):2187–2217, 2006.

F Jay Breidt, Nuno Crato, and Pedro De Lima. The detection and estimation of long memory in stochastic volatility. *Journal of econometrics*, 83(1):325–348, 1998.

Jonathan Brogaard. High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, page 66, 2010.

Peter Carr and Liuren Wu. Time-changed lévy processes and option pricing. *Journal of Financial Economics*, 71(1):113–141, 2004.

Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

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

Mikhail Chernov, A Ronald Gallant, Eric Ghysels, and George Tauchen. Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1):225–257, 2003.

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

Ray Yeutien Chou, Hengchih Chou, and Nathan Liu. Range volatility models and their applications in finance. In *Handbook of Quantitative Finance and Risk Management*, pages 1273–1281. Springer, 2010.

Peter K Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, pages 135–155, 1973.

Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

Fabienne Comte, Laure Coutin, and Éric Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.

Valentina Corradi and Walter Distaso. Semi-parametric comparison of stochastic volatility models using realized measures. *The Review of Economic Studies*, 73 (3):635–667, 2006.

Arnaud Doucet, Nando De Freitas, and NJ Gordon. Sequential monte carlo methods in practice. series statistics for engineering and information science, 2001.

Feike C Drost and Theo E Nijman. Temporal aggregation of garch processes. *Econometrica: Journal of the Econometric Society*, pages 909–927, 1993.

Michael Dueker and Christopher J Neely. Can markov switching models predict excess foreign exchange returns? *Journal of Banking & Finance*, 31(2):279–296, 2007.

Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

Robert F Engle and Jeffrey R Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162, 1998.

Bjørn Eraker. Mcmc analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, 19(2):177–191, 2001.

Eugene F Fama. The behavior of stock-market prices. *Journal of business*, pages 34–105, 1965.

David Freedman. *Brownian motion and diffusion*. Springer, 1971.

Mark Iosifovich Freidlin. *Functional Integration and Partial Differential Equations.(AM-109)*, volume 109. Princeton university press, 1985.

Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.

Hélyette Geman. From measure changes to time changes in asset pricing. *Journal of Banking & Finance*, 29(11):2701–2722, 2005.

Peter R Hansen and Asger Lunde. A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, 20(7): 873–889, 2005.

Peter R Hansen and Asger Lunde. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161, 2006.

Peter Reinhard Hansen, Zhuo Huang, and Howard Howan Shek. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906, 2012.

Andrew C Harvey. Long memory in stochastic volatility. *Forecasting volatility in the financial markets*, page 307, 2002.

Espen Gaarder Haug and Nassim Nicholas Taleb. Option traders use (very) sophisticated heuristics, never the black–scholes–merton formula. *Journal of Economic Behavior & Organization*, 77(2):97–106, 2011.

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.

Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2): 327–343, 1993.

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.

Xin Huang and George Tauchen. The relative contribution of jumps to total price variance. *Journal of financial econometrics*, 3(4):456–499, 2005.

John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300, 1987.

John C Hull. *Options, futures, and other derivatives.* Pearson Education India, 1999.

Eunju Hwang, Dong Wan Shin, et al. Stationary bootstrapping realized volatility under market microstructure noise. *Electronic Journal of Statistics*, 7:2032–2053, 2013.

Tsunehiro Ishihara and Yasuhiro Omori. Multivariate stochastic volatility model with cross leverage. In *Proceedings of COMPSTAT'2010*, pages 315–323. Springer, 2010.

Eric Jacquier, Nicholas G Polson, and Peter E Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87, 2002.

Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393, 1998.

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.

Peter Lax. *Linear Algebra and Its Applications.* John Wiley & Sons, 2nd edition, 2007.

Peter Lildholdt. *Estimation of GARCH models based on open, close, high, and low prices.* CAF, Centre for Analytical Finance, 2002.

Jane Liu and Mike West. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, pages 197–223. Springer, 2001.

Benoit Mandelbrot. The variation of some other speculative prices. *Journal of Business*, pages 393–413, 1967.

Worapree Maneesoonthorn, Catherine S Forbes, and Gael M Martin. Inference on self-exciting jumps in prices and volatility using high frequency measures. *arXiv preprint arXiv:1401.3911*, 2014.

Angelo Melino and Stuart M Turnbull. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1):239–265, 1990.

Robert C Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

Teruo Nakatsuma. Bayesian analysis of arma–garch models: A markov chain sampling approach. *Journal of Econometrics*, 95(1):57–69, 2000.

Daniel B Nelson. Arch models as diffusion approximations. *Journal of econometrics*, 45(1):7–38, 1990.

Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

Daniel B Nelson. Asymptotic filtering theory for multivariate arch models, 1994.

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.

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007.

Christian Pigorsch, Uta Pigorsch, and Ivaylo Popov. Volatility estimation based on high-frequency data. In *Handbook of Computational Finance*, pages 335–369. Springer, 2012.

Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

José M Quintana and Mike West. An analysis of international exchange rates using multivariate dlm's. *The Statistician*, pages 275–281, 1987.

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.

LCG Rogers, Fanyin Zhou, et al. Estimating correlation from high, low, opening and closing prices. *The Annals of Applied Probability*, 18(2):813–823, 2008.

Esther Ruiz. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, 63(1):289–306, 1994.

Jeffrey Russell and Robert F Engle. Econometric analysis of discrete-valued irregularly-spaced financial transactions data using a new auoregressive conditional multinomial model. 1998.

Tina Hviid Rydberg and Neil Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1): 2–25, 2003.

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.

Sandro Salsa. *Partial differential equations in action: from modelling to theory*, volume 99. Springer, 2016.

Gleb Sandmann and Siem Jan Koopman. Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics*, 87(2):271–301, 1998.

Vladimir Viktorovich Shaidurov. *Multigrid methods for finite elements*, volume 318. Springer Science & Business Media, 2013.

Neil Shephard. *Stochastic Volatility: Selected Readings.* Oxford University Press, UK, 2005.

Shinichiro Shirota, Takayuki Hizu, and Yasuhiro Omori. Realized stochastic volatility with leverage and long memory. *Computational Statistics & Data Analysis*, 76:618–641, 2014.

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.

Hans R Stoll. Presidential address: friction. *The Journal of Finance*, 55(4): 1479–1514, 2000.

Gilbert Strang and George J Fix. *An analysis of the finite element method*, volume 212. Prentice-hall Englewood Cliffs, NJ, 1973.

Makoto Takahashi, Yasuhiro Omori, and Toshiaki Watanabe. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis*, 53(6):2404–2426, 2009.

Viktor Todorov. Econometric analysis of jump-driven stochastic volatility models. *Journal of Econometrics*, 160(1):12–21, 2011.

Harald Uhlig. Bayesian vector autoregressions with stochastic volatility. *Econometrica: Journal of the Econometric Society*, pages 59–73, 1997.

JH Venter and PJ de Jongh. Extended stochastic volatility models incorporating realised measures. *Computational Statistics & Data Analysis*, 2012.

Ioannis D Vrontos, Petros Dellaportas, and Dimitris N Politis. Full bayesian inference for garch and egarch models. *Journal of Business & Economic Statistics*, 18(2):187–198, 2000.

RM Wilcox. Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, 8(4):962–982, 1967.

Jun Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127(2):165–178, 2005.

Jun Yu and Renate Meyer. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25(2-3):361–384, 2006.

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

Lan Zhang, Per A Mykland, and Yacine Aït-Sahalia. A tale of two time scales. *Journal of the American Statistical Association*, 100(472), 2005.

Bin Zhou. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52, 1996.

Gilles Zumbach. The pitfalls in fitting garch (1, 1) processes. In *Advances in Quantitative Asset Management*, pages 179–200. Springer, 2000.