Master of Science in Datascience

Minor-Dissertation: Final

Correlation emergence in two coupled limit order books in the fluid limit

Dominic Bauer

Supervisor: A/Prof. T. Gebbie

STA5093W

Department of Statistical Sciences,

University of Cape Town, Rondebosch

2023

Abstract

We use random walks to simulate the fluid limit of two coupled diffusive limit order books to model correlation emergence. The model implements the arrival, cancellation and diffusion of orders coupled by a pairs trader profiting from the mean-reversion between the two order-books in the fluid limit for a Lit order book with vanishing boundary conditions and order volume conservation. We are able to demonstrate the recovery of an Epps effect from this. We show how various stylised facts depend on the model parameters and the numerical scheme and discuss various strengths and weaknesses of the approach. We demonstrate how the Epps effect depends on different choices of time and price discretisation and show how an Epps effect can emerge without recourse to market microstructure effects.

Keywords: online learning, technical analysis, portfolio selection, backtesting, overfitting, in-sample, out-of-sample, Johannesburg Stock Exchange

Code: https://github.com/DominicGBauer/InteractingLOBs.jl

Contents

1 Introduction

Financial markets operate through intricate systems known as order books, wherein two prominent concepts, namely lit and latent order books, play pivotal roles. Lit order books provide an environment where all market participants can readily view and engage with displayed orders (O'Hara, 2015). This transparency not only fosters liquidity but also enables traders to promptly identify available orders, thereby contributing to the creation of a robust and efficient market.

A distinctive attribute of lit order books is their immediate execution mechanism. Trades within lit order books transpire swiftly at the best available price, aligning with the price-time priority principle (O'Hara, 2015). This characteristic makes lit order books particularly well-suited for investors seeking expeditious and transparent executions such as retail investors.

In contrast, latent order books (described by Tóth, Lempérière, Deremble, de Lataillade, Kockelkoren and Bouchaud (2011)) encompass hidden or non-displayed orders, providing a realm of confidentiality for traders. Frequently associated with dark pools (alternative trading systems designed for executing large trades away from public visibility (Budish, Cramton and Shim, 2015)), latent order books offer inherent confidentiality. This secretive nature of latent order books contributes to a reduced market impact for substantial trades, allowing market participants to execute sizeable transactions with minimised visibility and, consequently, mitigated potential influence on market prices (Menkveld, 2013).

Donier, Bonart, Mastromatteo and Bouchaud (2015) introduced the idea of using a reaction-diffusion model of the latent order book. Gant (2022a) implemented a numerical solution using the stochastic finite difference method (Angstmann, Henry, Jacobs and McGann, 2016; Angstmann, Donnelly, Henry, Jacobs, Langlands and Nichols, 2016) to solve for this model. The model was then extended by Diana and Gebbie (2023a), who introduced the idea of using two coupled order books to further improve the model.

We will use this improved model to simulate two coupled lit order books to verify if the solution proposed in Diana and Gebbie (2023a) has the Epps effect (Epps, 1979) as an emergent property as this has been observed in various empirical studies (Mastromatteo, Marsili and Zoi, 2010) of financial markets.

The structure of the dissertation is as follows. Section 2 explains the numerical scheme used to implement the model practically. Section 3 explains the coupled order book model used to produce realistic price paths. Section 4 introduces the Epps effect and the non-uniform fast Fourier transform used to estimate the correlation between the price paths generated from the model. It further verifies if the proposed coupled order book model has the Epps effect as an emergent property and explores the effect of changing parameters on the Epps effect. Section 5 is where we attempt to calibrate the model by using empirical data to optimise our parameters using the method of moments with simulated minimum distance. We further investigate the stylised facts generated by the calibrated model and compare them with empirical stylised facts. Section 6 is where we summarise and conclude our results.

Briefly, we will show that the model proposed by Diana and Gebbie (2023a) produces an Epps effect when using the calibrated model parameters with non-uniform sampling. This Epps effect can be tuned to remove some of the numerical artefacts and is found without the need for market microstructure noise models relative to a preferred latent model. The Epps effect will be entirely due to discretisation and sampling. However we will highlight that the numerical artefacts in the simulation that seem to be due to the discrete sampling while enforcing the diffusive nature of

randomness needs further investigation.

2 Numerical Scheme used for Coupled Order Book

2.1 The simple update equation

From the discrete master equation formulation of the reaction-diffusion Equation [22] at time t_n on a time and price grid we can consider the overall density of orders by finding: the orders that arrive, orders that are removed, the orders that move from different positions on the surrounding lattice to the current price x_i due to random effects over the time increment Δt, and those that have remained from time t_{n-1} and self interact, also from time t_n to time t_n at price x_i. This then provides the order density φ_n^i at time t_n and price x_i. Here the right, left and self jumps transition rates back to state i from price states $i+j$ for $j \in \{-1, 0, +1\}$ can be given as probabilities on the lattice $\lambda_n^{i|i+j}$:

$$\varphi_n^i = C_n^i - A_{n-1}^i \varphi_{n-1}^i + \sum_{j=-1}^{j=+1} \lambda_{n-1}^{i|i+j} \varphi_{n-1}^{i+j}.$$

Where the arriving orders added over the time increment Δt is C_n^i and the number of orders removed over the time increment Δt is A_{n-1}^i. Then the orders that survive over that time interval can be written as the complement of a survival function as: $A_{n-1}^i = 1 - \theta_{n-1,n}^i$. These are orders that survive removal at time t_{n-1} until time t_n. The master equation then becomes:

$$\varphi_n^i = C_n^i + \theta_{n-1,n}^i \varphi_{n-1}^i + \sum_{j=-1}^{j=+1} \lambda_{n-1}^{i|i+j} \varphi_{n-1}^{i+j} - \varphi_{n-1}^i.$$

We pick the survival function in terms of a Poisson removal process: $\theta_{n-1,n}^i = e^{-\nu \Delta t}$ and write the arriving orders as a source density function $s(x_i, t_{n-1})$ evaluated at time t_{n-1}.

$$\varphi_n^i = s(x_i, t_{n-1})\Delta t + e^{-\nu \Delta t} \varphi_{n-1}^i + \sum_{j=-1}^{j=+1} \lambda_{n-1}^{i|i+j} \varphi_{n-1}^{i+j} - \varphi_{n-1}^i.$$

The right, left and self jumps transition rates back to state i from price states $i+j$ for $j \in [-1, 0, +1]$ can be given as probabilities on the lattice: $\lambda_n^{i|i+j} = P_n^{(-j)}$ where $P_n^{(+1)}$, $P_n^{(0)}$ and $P_n^{(-1)}$ are, respectively, the probabilities of right, left and self jump transitions over the time increment $t_n - t_{n-1}$ implemented at time t_n:

$$\varphi_n^i = s(x_i, t_{n-1})\Delta t + e^{-\nu \Delta t} \varphi_{n-1}^i + P_{n-1}^{(+1)} \varphi_{n-1}^{i-1} + P_{n-1}^{(-1)} \varphi_{n-1}^{i+1} + (P_{n-1}^{(0)} - 1)\varphi_{n-1}^i. \quad (1)$$

To include a discrete driving force $F_n = f(t_n)$ from $f = -\partial_x V$ we update the probability of left, right and self jumps respectively:

$$P_n^{(+1)} = \tfrac{1}{2}(r - F_n), \quad P_n^{(-1)} = \tfrac{1}{2}(r + F_n), \text{ and } P_n^{(0)} = 1 - r; \quad r \in [0, 1].$$

These are substituted into Equation [1]:

$$\varphi_n^i = \tfrac{1}{2}(r - F_{n-1})\varphi_{n-1}^{i-1} + \tfrac{1}{2}(r + F_{n-1})\varphi_{n-1}^{i+1} - (r - e^{-\nu \Delta t})\varphi_{n-1}^i + s(x_i, t_{n-1})\Delta t. \quad (2)$$

2.2 The reaction-diffusion PDE

We start with the master equation from Equation [1]:

$$\varphi_n^i = s(x_i, t_{n-1})\Delta t + e^{-\nu\Delta t}\varphi_{n-1}^i + P_{n-1}^{(+1)}\varphi_{n-1}^{i-1} + P_{n-1}^{(-1)}\varphi_{n-1}^{i+1} + (P_{n-1}^{(0)} - 1)\varphi_{n-1}^i. \quad (3)$$

We can write the values of the lattice points $x_i = i\Delta t$ and $t_n = n\Delta t$ in terms of approximating functions: $P_n^{(-1)} = p_\Delta^-(x_i, t_n)$, $P_n^{(+1)} = p_\Delta^+(x_i, t_n)$, $P_n^{(0)} = p_\Delta(x_i, t_n)$, $\varphi_n^i = \varphi_\Delta(x_i, t_n)$, and $S_n^i = s_\Delta(x_i, t_n)$:

$$\begin{aligned}\varphi_\Delta(x_i, t_n) = s_\Delta(x_i, t_{n-1})\Delta t + e^{-\nu\Delta t}\varphi_\Delta(x_i, t_{n-1}) \\ + p_\Delta^+(x_i, t_{n-1})\varphi_\Delta(x_{i-1}, t_{n-1}) \\ + p_\Delta^-(x_i, t_{n-1})\varphi_\Delta(x_{i+1}, t_{n-1}) \\ + (p_\Delta(x_i, t_{n-1}) - 1)\varphi_\Delta(x_i, t_{n-1}).\end{aligned} \quad (4)$$

We substitute $x = i\Delta x$ and $t = n\Delta t$:

$$\begin{aligned}\varphi_\Delta(x, t) = s_\Delta(x, t - \Delta t)\Delta t + e^{-\nu\Delta t}\varphi_\Delta(x, t - \Delta t) \\ + p_\Delta^+(x, t - \Delta t)\varphi_\Delta(x - \Delta x, t - \Delta t) \\ + p_\Delta^-(x, t - \Delta t)\varphi_\Delta(x + \Delta x, t - \Delta t) \\ + (p_\Delta(x, t - \Delta t) - 1)\varphi_\Delta(x, t - \Delta t).\end{aligned} \quad (5)$$

The diffusion probabilities sum to 1:

$$1 = p_\Delta^-(x_i, t_n) + p_\Delta^+(x_i, t_n) + p_\Delta(x_i, t_n). \quad (6)$$

We define a stochastic force function in terms of the diffusion jump transitions (See Section 2.7) using an inverse temperature β and some thermodynamic constant C:

$$f_\Delta(x, t) = \frac{C}{\beta\Delta x}\left(p_\Delta^+(x, t) - p_\Delta^-(x, t)\right). \quad (7)$$

Here a stochastic force opposes the left and right diffusion jumps when they occur. We normalise the stochastic force jump probabilities: $C(p_\Delta^+ + p_\Delta^-) = 1$ where the probability of not jumping due to the diffusion is $(1 - r)$ so that the probability of left and right diffusion jumps are $\frac{r}{2}$ and $C = \frac{1}{r}$ where r is non-zero. If we then add and subtract $p_\Delta^-(x, t)$ from probability normalisation equation for the diffusion (Equation [6]), and substitute for the approximating function for the stochastic force on the lattice to get:

$$p_\Delta^\pm(x, t) = \tfrac{1}{2}\left(1 - p_\Delta(x, t) \pm r\beta\Delta x f_\Delta\right). \quad (8)$$

Use (8) in (5), Taylor expand the exponential, and cancel terms:

$$\begin{aligned}\varphi_\Delta(x, t) = s_\Delta(x, t - \Delta t)\Delta t - \nu\varphi_\Delta(x - \Delta x, t - \Delta t)\Delta t \\ + \tfrac{1}{2}(1 - p_\Delta + r\beta\Delta x f_\Delta)\varphi_\Delta(x - \Delta x, t - \Delta t) \\ + \tfrac{1}{2}(1 - p_\Delta - r\beta\Delta x f_\Delta)\varphi_\Delta(x + \Delta x, t - \Delta t) \\ + p_\Delta\varphi_\Delta(x, t - \Delta t).\end{aligned} \quad (9)$$

Subtract $\varphi_\Delta(x, t-\Delta t)$ and divide by Δt, on both sides, use that $p_\Delta = 1 - r$ for some jump probability r, and then take the lattice grid limits $\Delta t \to 0$ and $\Delta x \to 0$ to find:

$$\begin{aligned}
&\lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} \frac{\varphi_\Delta(x,t) - \varphi_\Delta(x,t-\Delta t)}{\Delta t} \\
&= \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} s_\Delta(x,t-\Delta t) - \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} \nu\varphi_\Delta(x-\Delta x,t-\Delta t) \\
&+ \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} \frac{r}{2}\frac{\Delta x^2}{\Delta t}\left[\frac{\varphi_\Delta(x-\Delta x,t-\Delta t) - 2\varphi_\Delta(x,t-\Delta t) + \varphi_\Delta(x+\Delta x,t-\Delta t)}{\Delta x^2}\right] \\
&- \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} 2\beta f_\Delta \frac{r}{2}\frac{\Delta x^2}{\Delta t}\left[\frac{\varphi_\Delta(x+\Delta x,t-\Delta t) - \varphi_\Delta(x-\Delta x,t-\Delta t)}{2\Delta x}\right]
\end{aligned} \tag{10}$$

Assume that the following diffusion limit exists:

$$D = \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} \frac{r}{2}\frac{\Delta x^2}{\Delta t}. \tag{11}$$

We take the diffusion limit, use the definitions of first[1] and second derivatives[2], and that for some approximating function (say) $y_\Delta(x,t)$ (a function which takes the limits over the lattice) on the lattice that:

$$y(x,t) = \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} y_\Delta, \tag{12}$$

where for another approximating function (say) z_Δ that:

$$y(x,t)z(x,t) = \lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} y_\Delta z_\Delta. \tag{13}$$

We can then find that the update equation is equivalent to the reaction-diffusion partial differential equation:

$$\partial_t\varphi = +s(x,t) - \nu\varphi + D\partial_{xx}\varphi - 2\beta D f\partial_x\varphi. \tag{14}$$

We can define a forcing function in terms of a potential $V(x,t) = -\frac{V_t x}{2D\beta}$ using that $f(x,t) = -\partial_x V(x,t)$:

$$\partial_t\varphi = +s(x,t) - \nu\varphi + D\partial_{xx}\varphi + V_t\partial_x\varphi. \tag{15}$$

2.3 Boundary conditions

The boundary conditions are defined at the highest and lowest price points denoted with the lowest price as $x = x_0 = 0$, and the highest price as $x = x_M$ on a lattice of $1 + M$ points.

We considered three choices of boundaries: 1.) *periodic boundaries*, 2.) using *ghost points* to enforce a zero-flux boundary (Angstmann et al., 2016), and 3.) *reflective boundaries* (Benzaquen and Bouchaud, 2018). We chose to use "ghost points" at $i = 0$ and $i = M + 1$ determined from the zero-flux Neumann boundary conditions (Mazumder, 2016) as this proved more convenient for our use case.

[1] $y_x = \lim_{\Delta x\to 0} \frac{1}{2\Delta x}(y(x+\Delta x) - y(x-\Delta x))$
[2] $y_{xx} = \lim_{\Delta x\to 0} \frac{1}{\Delta x^2}(y(x+\Delta x) - 2y(x) + f(x-\Delta x))$

2.4 Ghost points

Using Neumann boundary conditions we can determine the values of the ghost points:

$$\varphi_n^{-1} = \varphi_n^0 - \Delta x \varphi_x(x, t_n)|_{x=x_0},$$
$$\varphi_n^{M+1} = \varphi_n^M + \Delta x \varphi_x(x, t_n)|_{x=x_M}.$$

These can be further reduced using the zero-flux condition where $\varphi_x(x_0, t) = 0 = \varphi_x(x_M, t)$ or more intuitively where the boundaries are flat. Near the minimum spatial state, we can substitute for the ghost points:

$$\varphi_n^0 = \tfrac{1}{2} r \varphi_{n-1}^{-1} + \tfrac{1}{2} r \varphi_{n-1}^1 - (1 - e^{-\nu \Delta t}) \varphi_{n-1}^0 + s(x_0, t_{n-1}) \Delta t.$$

Similarly, near the maximum state we have:

$$\varphi_n^M = \tfrac{1}{2} r \varphi_{n-1}^{M-1} + \tfrac{1}{2} r \varphi_{n-1}^{M+1} - (r - e^{-\nu \Delta t}) \varphi_{n-1}^M + s(x_M, t_{n-1}) \Delta t.$$

2.5 Initial conditions

The first step requires solving for the initial conditions of the SPDE. We define the initial conditions $\varphi(x, 0)$ by first approximating the derivatives with the finite difference method: $\varphi_t(x_i, 0) \approx 0$, $\varphi_x(x_i, 0) \approx \frac{1}{2\Delta x}(\varphi_0^{i+1} - \varphi_0^{i-1})$, and $\varphi_{xx}(x_i, 0) \approx \frac{1}{\Delta x^2}(\varphi_0^{i+1} - 2\varphi_0^i + \varphi_0^{i-1})$. We can then rewrite the reaction-diffusion equation using the initial conditions to find:

$$-s_0^i = \tfrac{D}{\Delta x^2}(\varphi_0^{i+1} - 2\varphi_0^i + \varphi_0^{i-1}) - \frac{V_0}{2\Delta x}(\varphi_0^{i+1} - \varphi_0^{i-1}) - \nu \varphi_0^i.$$

Now we can group term and re-write this as a $1 \times (M+1)$ matrix equation: $\boldsymbol{s} = A\boldsymbol{\varphi}$. Using the zero-flux Neumann boundary condition we modify the first and last rows of A. Then with a linear solver we can then find the state vector of initial conditions with components: φ_0^i (Angstmann et al., 2016)

$$A = \begin{bmatrix} [2] - \frac{2D}{\Delta x^2} - \nu & \frac{2D}{\Delta x^2} & & & & \\ \frac{D}{\Delta x^2} + \frac{V_0}{2\Delta x} & -\frac{2D}{\Delta x^2} - \nu & \frac{D}{\Delta x^2} - \frac{V_0}{2\Delta x} & & & \\ & \frac{D}{\Delta x^2} + \frac{V_0}{2\Delta x} & -\frac{2D}{\Delta x^2} - \nu & \frac{D}{\Delta x^2} - \frac{V_0}{2\Delta x} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{D}{\Delta x^2} + \frac{V_0}{2\Delta x} & -\frac{2D}{\Delta x^2} - \nu & \frac{D}{\Delta x^2} - \frac{V_0}{2\Delta x} \\ & & & & \frac{2D}{\Delta x^2} & -\frac{2D}{\Delta x^2} - \nu \end{bmatrix}.$$

2.6 Grid dynamics and stability

The Dirichlet boundary conditions can be given at x_0 using the initial condition derived by solving the SPDE via finite difference (see Section 2.5). However, for numerical stability, the density must be conserved such that $\sum_{i=1}^M \varphi_{i,n} = \sum_{i=1}^M \varphi_{i,n-1}$. If the initial density is finite the conservation of density can ensure numerical stability (Angstmann et al., 2016).

To ensure numerical stability and that the granularity of the mesh can reasonably capture the velocity of density of order volume propagation we need the distance between the approximation and the numerical scheme to be bounded for all n time steps after the application of initial and boundary conditions (Angstmann et al., 2016):

$$\sum_{i=1}^{M} |\varphi_n^i - \varphi(x_0 + i\Delta x, n\Delta t)| \leq C, \quad C \in \mathbb{R}^+. \tag{16}$$

This requires that the maximum probability of left and right jumps correspond to the maximum velocity allowed by the mesh. However, the presence of viscosity ensures the maximum density necessarily occurs at $t = 0$ when initial conditions are computed (Angstmann et al., 2016).

2.7 Transition probabilities

The prices are assumed to evolve on a discrete lattice (x_i, t_n) with price state $x_i = i\Delta x$ for price separation Δx and times $t_n = n\Delta t$ for some temporal separation Δt. Again, use state-dependent transition probabilities from the i-th lattice point at time t_n to construct a master equation. We first consider three transition types on this lattice: a right-transition $P_{i,n}^{(+1)}$, a left-transition $P_{i,n}^{(-1)}$, a self-transition $P_{i,n}^{(0)}$. Additionally, there are cancellations (order removals) at rate $a(x,t)$, and new particles can arrive and be deposited on the lattice according to a source term $c(x_i, t_n)$.

Jump probabilities can be associated with continuous functions sampled at discrete points on such a lattice: $P_{i,n}$, and if an appropriate diffusion limit is taken (Angstmann et al., 2016; Angstmann et al., 2016) for diffusion constant $D = \lim_{\Delta t, \Delta x \to 0} \frac{r}{2} \frac{\Delta x^2}{\Delta t}$ a discrete numerical scheme to simulate the limit-order book: $\varphi_{i,n} \approx \varphi(i\Delta x, n\Delta t)$ can then be used to find sample paths of mid-prices $p(t)$ as the prices at discrete time t_n where $\varphi(x,t) = 0$.

The transition probabilities between states can be represented using Boltzmann potentials $V(x_i, t_n)$ to describe the information arrival in terms of a stochastic driving force V_t: $V(x_i, t_n) = -\frac{V_{t_n} x_i}{2D\beta}$ *e.g.* i.i.d. white noise with variance σ_f^2. The inverse temperature is β which in our context could be seen as an approximation for market volatility. This is useful because the state and temporal lattice differences can be defined in terms of the minimum and maximum prices, $x_0 = 0$ and x_M, for some M, and the diffusion parameter D: $\Delta x = \frac{x_0 - x_M}{M}$ and $\Delta t = \frac{\Delta x^2}{2D}$ where $i \in [0, M]$.

The Boltzmann potential then represents the energy of the price state and transitions to price $x_i = i\Delta x$ for $i = 0, 1, 2, \ldots, M$. This can be used to compute the Boltzmann probability of the particle being in some price at some time. The jump probability is dependent on the Boltzmann potential of the destination state.

The jump probabilities for the left, self and right jump are respectively $j = -1, 0, +1$: $P_{i,n}^{(j)} = Z_{i,n}^{-1} e^{-\beta V(x_{i+j}, t_n)}$. The normalization factor $Z_{i,n}$ at price x_i at time t_n is the sum over all three allowable state transition probabilities for $i \in [1, M-1]$: $Z_{i,n} = \sum_{j=-1}^{+1} e^{-\beta V(x_{i+j}, t_n)}$. The normalisations ensure that transition probabilities behave like probabilities on $[0, 1]$.

Further simplification of the jump probabilities is possible due to the design of the price discretization: $x_{i+1} = x_i + \Delta x$ and $x_{i-1} = x_i - \Delta x$. By substituting in the jump probability equations, we can factor out a common term and remove the jump probabilities' dependence on x. This simplifies the jump probability normalisation to $Z_{i,n} = \exp(\frac{V_{t_n} \Delta x}{2D}) + \exp(-\frac{V_{t_n} \Delta x}{2D}) + 1$ and the individual jump probabilities become: $P_n^{(j)} = Z_{i,n}^{-1} \exp(\frac{j V_{t_n} \Delta x}{2D})$.

2.8 Stochastic time update equation

Consider the final update Equation [2]. Here Δt and Δx are constant during the simulation. This defines a uniform lattice. Now let us rather suppose that we generate a sequence of changes in time, which we will denote $\{\Delta t_m\}_{m=1}^{M}$. The simulation starts at t_0 and continues until time t_M. These time steps are no longer uniform. We can do this because the diffusion is built on a random walk process as are the stochastic forces and these are all price independent probabilities, but there can be time dependence for the stochastic force. Using Equation [31] for the lattice sizes with fixed values for D and r we generate for each n, a corresponding lattice spacing sequence $\{\Delta x_k\}_{k=1}^{M}$ were we denote Δx to be the average grid width: $\Delta x = \mathbb{E}[\Delta x_k]$.

A choice of Δt_{n-1} and thus Δx_{n-1}, determines the price gap from where the left and right jumps to price x_{n-1}^i will occur at time t_{n-1} to time t_n. Here these points are: $x_{n-1}^{i\pm1} = x_i \pm \Delta x_{n-1}$. We now introduce a background lattice (x_i, t_n) which serves as a reference framework to track the evolution of prices over time, and a jump lattice associated with the jumps: x_n^i. To proceed with the simulation we first generate the background lattice, this will have non-uniformly sampled times but uniformly sampled prices. We use the average grid size Δx to generate the background lattice: (x_i, t_n). We assume there is some approximation function $\varphi_\Delta(x,t)$ for the densities on the background lattice: $\varphi_n^i = \varphi_\Delta(x_i, t_n)$. Then we can find the prices from which the right and left jumps will occur that are consistent with the diffusion to the order of the approximations from the background:

$$\hat{\varphi}_n^{i\pm1} = \varphi_\Delta(x_n^{i\pm1}, t_n) = \varphi_\Delta(x_i \pm \Delta x_n, t_n). \tag{17}$$

The probabilities of left and right jumps do not depend on the sequence $\{\Delta x_n\}_{n=1}^{M}$. They do depend on the most recent entry at t_{n-1}. Using this we can then rewrite [2]:

$$\begin{aligned}\varphi_n^i = {} & \tfrac{1}{2}(r - F_{n-1})\hat{\varphi}_{n-1}^{i-1} + \tfrac{1}{2}(r + F_{n-1})\hat{\varphi}_{n-1}^{i+1} \\ & - (r - e^{-\nu\Delta t_{n-1}})\varphi_{n-1}^i + s(x_i, t_{n-1})\Delta t_{n-1}.\end{aligned} \tag{18}$$

We can identify x_i in the background with x_n^i:

$$\hat{\varphi}_{n-1}^{i\pm1} \approx \varphi_\Delta(x_i \pm \Delta x_{n-1}, t_{n-1}). \tag{19}$$

At time t_{n-1} we use the background lattice point (x_i, t_n) with a first-order mid-point approximation to find the densities:

$$\hat{\varphi}_{n-1}^{i\pm1} \approx \varphi_{n-1}^{i\pm1} + \tfrac{(\Delta x - \Delta x_{n-1})}{2\Delta x}\left[\varphi_{n-1}^i - \varphi_{n-1}^{i\pm2}\right]. \tag{20}$$

Here when $\Delta x_{n-1} \to \Delta x$ then $\hat{\varphi}_{n-1}^{i\pm1} \to \varphi_{n-1}^{i\pm1}$ and the background and the lattice coincide. This only holds for small Δx_n.

However, in order to accommodate $\Delta x_{n-1} > \Delta x$ on the jump lattice we first search the background lattice for some point x_k such that $x_k \le x_{n-1}^{i\pm1} \le x_{k+1}$ and then use a first order mid-point approximation to find the densities:

$$\hat{\varphi}_{n-1}^{i\pm1} \approx \varphi_{n-1}^{k} + \tfrac{(x_{n-1}^{i\pm1} - x_k)}{2\Delta x}\left[\varphi_{n-1}^{k+1} - \varphi_{n-1}^{k-1}\right]. \tag{21}$$

We can then use that $\Delta t_{n-1} \sim \mathrm{Exp}(\lambda)$ where the intensity λ can be chosen such that $\Delta t = \mathbb{E}[\Delta t_{n-1}] = \frac{1}{\lambda}$.

3 Coupled Order Book

3.1 Coupled partial differential equations

Following Diana and Gebbie (2023a) and using the extended implementation of Gant (2022b) we define two order book equations, $j \in \{1, 2\}$ in the presence of information shocks that can be synchronised (or remain unsynchronised) where the reaction-diffusion equation describing the j^{th} order book is:

$$\partial_t \varphi^{(j)} = D_j \partial_{xx} \varphi^{(j)} + V_t^{(j)} \partial_x \varphi^{(j)} - a^{(j)} \varphi^{(j)} + c^{(j)}. \tag{22}$$

Here j is the order book index so that, for example, D_1 and D_2 are the order diffusion rates for the two limit order books with random driving forces $V_t^{(1)}$ and $V_t^{(2)}$. We have that the trade prices for the j^{th} order book are:

$$p^{(j)}(t) = x \ \text{ s.t. } \ \varphi^{(j)}(x,t) = 0. \tag{23}$$

We assume that the annihilation rates are constants:

$$a^{(j)}(x,t) = \nu_j. \tag{24}$$

The creation term is separated further into a source terms $s^{(j)}$, coupling terms $\ell^{(j,k)}$, and shocks $\delta^{(j)}$ for the j^{th} assets order book:

$$c^{(j,k)}(x,t) = s^{(j)}(x,t) + \ell^{(j,k)}(x,t) + \delta^{(j)}(x,t). \tag{25}$$

The source terms are assumed to be either latent order book sources or lit order book sources. If there is an order book coupling from the j^{th} order book to (say) the k^{th} order book this will be carried by the coupling term $\ell^{(j,k)}$. If there is no coupling then we can drop the k index on the source term. Boundary condition considerations (to ensure that that the integral under $\varphi^{(i)}$ is constant) make it convenient that we will only consider the lit order book sources (Gant, 2022a; Diana and Gebbie, 2023a):

$$s^{(j)}(x,t) = -\lambda_j \mu_j (x - p^{(j)}(t)) e^{\mu_j \left(x - p^{(j)}(t)\right)^2} \tag{26}$$

3.1.1 Order book couplings

We use the pair-wise couplings proposed by Diana and Gebbie (2023a) and implemented in Diana and Gebbie (2023b) between the j^{th} and k^{th} order books:

$$\begin{aligned} \ell^{(j,k)}(x,t) &= G(x,t,p^{(j)}, \Delta p_{jk}) = G_j(x,t) \\ \ell^{(k,j)}(x,t) &= G(x,t,p^{(k)}, \Delta p_{kj}) = G_k(x,t) \end{aligned} \tag{27}$$

The coupling equations are a function of the difference in mid-prices of two order books where $p^{(j)}$ in $\varphi^{(j)}$ is above $p^{(k)}$ (in $\varphi^{(k)}$), then more bids are placed above the mid-price in $\varphi^{(j)}$ to push the mid-price down to $\varphi^{(k)}$'s mid-price. To achieve this we define:

$$g_j(x,t) = -\lambda_j \mu_j x e^{\mu_j x^2}, \tag{28}$$

where λ_j and μ_j are constants specific to the j^{th} order book, and Δp_{jk} is the difference between the mid-prices of the two order books: $\Delta p_{jk} = p^{(j)} - p^{(k)}$:

$$G_j = \begin{cases} g_j\left(x - p^{(j)}(t)\right)\ \Delta p_{jk} & , x > p^{(j)}(t), \Delta p_{jk} > 0 \\ g_j\left(\frac{j}{\Delta p_{jk}}(x - p^{(j)}(t))\right) & , x \leq p^{(j)}(t), \Delta p_{jk} > 0 \\ g_j\left(x - p^{(j)}(t)\right)\ \Delta p_{jk} & , x \leq p^{(j)}(t), \Delta p_{jk} \leq 0 \\ g_j\left(\frac{1}{\Delta p_{jk}}(x - p^{(j)}(t))\right) & , x > p^{(j)}(t), \Delta p_{jk} \leq 0. \end{cases}$$

This has the interpretation of an external agent (such as a pairs trader) observing the system, and buying (or selling) the one asset according to whether the mid-price of the other asset is as described above. In this way the system can be generalised to many assets being traded by pair-traders. Here we will focus on only two assets.

3.1.2 Order book shocks

Order book shocks are used to estimate the price impact (Diana and Gebbie, 2023a). These involve introducing a shock based of size Q and measuring the change in price as a result.

3.1.3 Information shocks

The driving force $V_t^{(j)}$ is taken to be a Brownian Motion. When the information shocks are synchronised then there is only one source for the stochastic force potential V_t and we drop the k index.

3.2 Coupled update equations

The coupled simple diffusion equation given in Equation [22] can be numerically solved using an update equation (see Section 2.1) which can be extended to include non-uniformly sampled times (see Section 2.8) and is then given in Equation [29] for a set of coupled order books:

$$\varphi_n^{i(j)} = \tfrac{1}{2}(r - F_{n-1}^{(j)})\hat{\varphi}_{n-1}^{i-1(j)} + \tfrac{1}{2}(r + F_{n-1}^{(j)})\hat{\varphi}_{n-1}^{i+1(j)} - (r - e^{-\nu_j \Delta t_{n-1}})\varphi_{n-1}^{i(j)} + c_{i,n-1}^{(j,k)}\Delta t_{n-1}. \tag{29}$$

3.2.1 Lattice parameters

As we assume that the diffusion limit exists in Equation [11]:

$$D = \frac{r}{2}\frac{\Delta x^2}{\Delta t} \tag{30}$$

We can then use this to set the lattice price grid increments in terms of the lattice time increments[3]:

$$\Delta x = \sqrt{\frac{2D}{r}\Delta t} \tag{31}$$

[3]Note that this implies an uncertainty principle: $\langle \Delta x \frac{\Delta x}{\Delta t} \rangle = \frac{2D}{r}$ and $\langle \Delta t \frac{\Delta x^2}{\Delta t^2} \rangle = \frac{2D}{r}$ that limits the minimum allowed grid combinations on the lattice. The price changes in terms of the velocity of the prices, and the time changes in terms of acceleration of the prices.

Here we first select the time increment Δt and then use this to set the price increments Δx. Similarly if we first select Δx then Δt is:

$$\Delta t = \frac{r}{2}\frac{\Delta x^2}{D} \tag{32}$$

To simulate non-uniformly sampled grids in time we then assume that the time changes are exponentially distributed: $\Delta t \sim \text{Exp}(\lambda_i t)$. Using this, we can have two different clocks, one for each of the two order books as defined by the two intensities, λ_1 and λ_2 (Diana and Gebbie, 2023a).

3.2.2 Limit order book parameters

The following parameters are used in the implementation by Diana and Gebbie (2023a,b):

Parameter	Description	Type	Value
L	System length	Fixed	200
M	Number of divisions	Fixed	400
r	Probability of self jump	Fixed	0.5
D	Diffusion rate	Free	0.5
ν	Cancellations rate	Free	14.0
γ	Fraction of derivative	Free	1.0
$p_{1,2}(0)$	Initial prices	Fixed	230.0
$\lambda_{1,2}$	Source terms intensities	Fixed	1.0
$\mu_{1,2}$	Source terms rates	Fixed	0.1
Δx	Change in x ($\frac{L}{M}$)	Fixed	0.5
Δt	Change in t (see Equation [32])	Free	0.0625

Table 1: The base model parameters used in the coupled order book model with accompanying description. Also shown is whether a value was fixed or free when the calibration of the model was conducted in Section 5. Δt is a free parameter due to it being a function of D which is a free parameter. These parameters are used in this section to plot figures 1, 2 & 3. They are also used in Section 4 to plot Figure 5 and we change the values of Δt and Δx for figures 6 & 7, respectively and the value of ν for Figure 8.

In Table 3.2.2 we provide the base parameters used in the model. These values were arbitrarily chosen as an initial starting point. L is the system length and M is the number of divisions within the system length. Both are used to determine the size of Δx where $\Delta x = \frac{L}{M}$. D is the diffusion rate described in Equation [30], r is the probability of a jump occurring (which would lead a change in price), ν is the rate at which orders in the order book are cancelled, γ is the fraction of the derivative (in Equation [30] we have shown the case where $\gamma = 1$), $p_{1,2}(0)$ are the initial prices for the two order books, $\lambda_{1,2}$ and $\mu_{1,2}$ are the source term variables defined in Equation [26] for the two order books.

3.3 Coupled order books visualisation

3.3.1 Price paths

In Figure 1 we demonstrate price paths generated using our model when considering a two-coupled order book case using the parameters from Table 3.2.2. The price paths

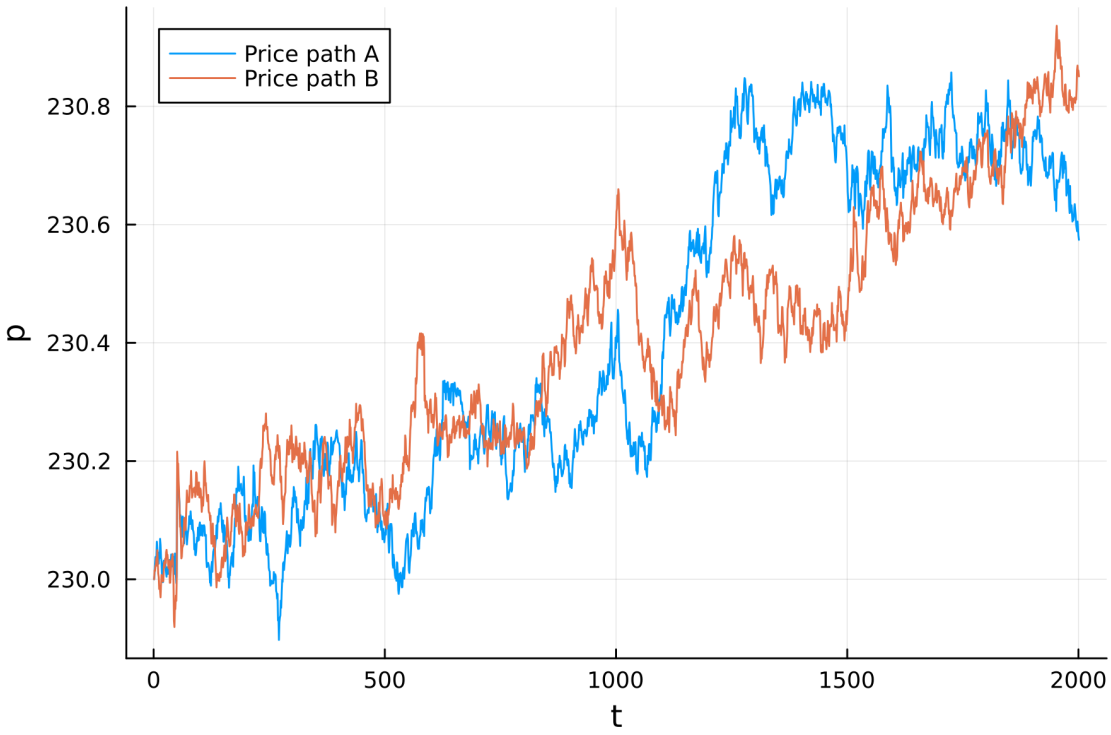
in the figures for order book A and B are blue and red, respectively. In Figure 1a we use the case of non-uniform Δt as described in Section 3.2.1 in our model. In Figure 1b we use the case of a uniform Δt (see Equation [32]) in our model.

3.3.2 Price dynamics

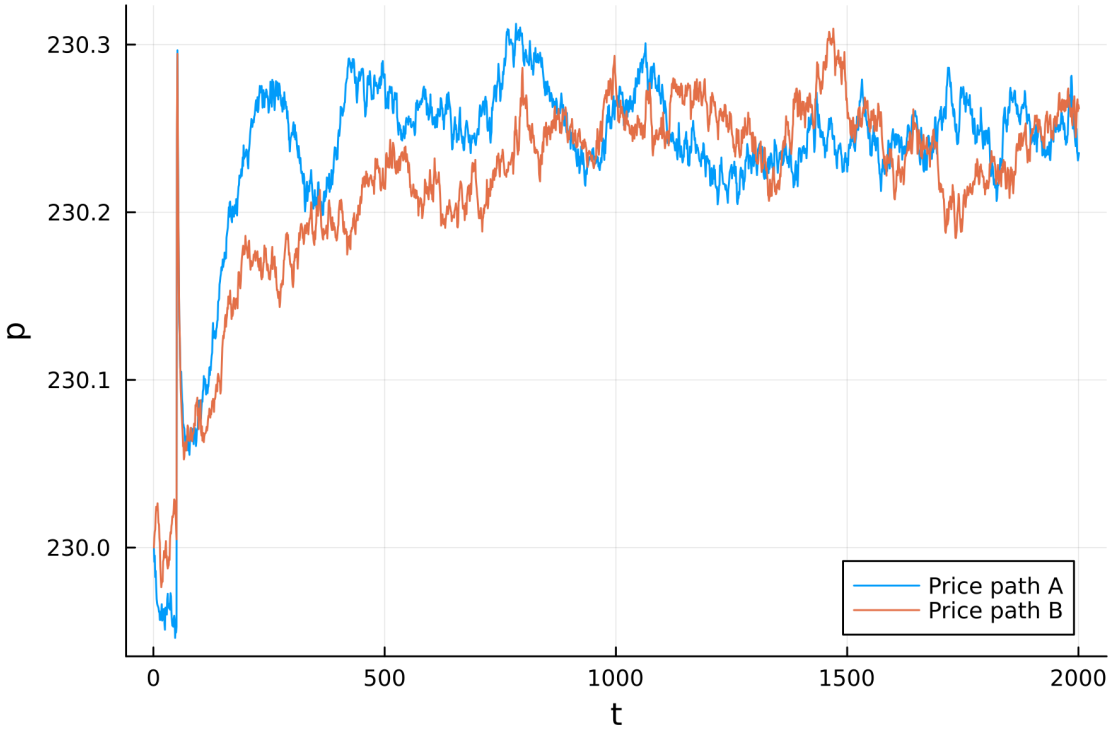
In figures 2 and 3 we showcase through snapshots of time increments Δt the effect an order book shock has on the mid-price p when applying a non-uniform Δt and uniform Δt to our model, respectively. In the subplots for both figures, blue represents the density of the order book we are modelling at the current time, green represents the orders which are about to be added via the source term, gold represents the orders which are about to be removed via the removal rate and pink represents an order book shock. The state of equilibrium in the system is where all the lines meet (i.e. the order density is 0) and is the mid-price $p.p$ and x_m are shown at the top right for each time increment Δt_m.

When looking at the non-uniform Δt case in Figure 2 we start with an initial order book shock as shown by the pink spike in Figure 2a. This results in an increase in density and a counteracting decrease in removals as shown in Figure 2b. This increase in density and decrease in removals persists in figures 2c, 2d and 2e at which point the shock is gone and the density and removals remain. Note that the removal density is smaller in magnitude to the density of the order book indicating that there is more willingness to buy than to sell and as such will result in a price increase to accommodate for this increased demand. We see this play out as the system seeks a new equilibrium by the increases in p as seen in figures 2f, 2g and 2h. Finally in Figure 2i the system reaches equilibrium at an increased mid-price $p = 230.19$.

Similarly, when looking at the uniform Δt case in Figure 3 we start with an initial order book shock as shown by the pink spike in Figure 3a. This results in an increase in density and a counteracting decrease in removals as shown in Figure 3b. This increase in density and decrease in removals persists in Figure 3c at which point the shock is gone and the order book density and removals remain. Note again that the removal density is smaller in size to the density of the order book indicating that there is more willingness to buy than to sell and as such will result in a price increase to accommodate for this increased demand. We see this play out as the system seeks a new equilibrium by the increases in p as seen in figures 3d and 3e. Finally in figures 3g, 3h and 3i we see the system reach equilibrium at an increased mid-price $p = 230.32$.

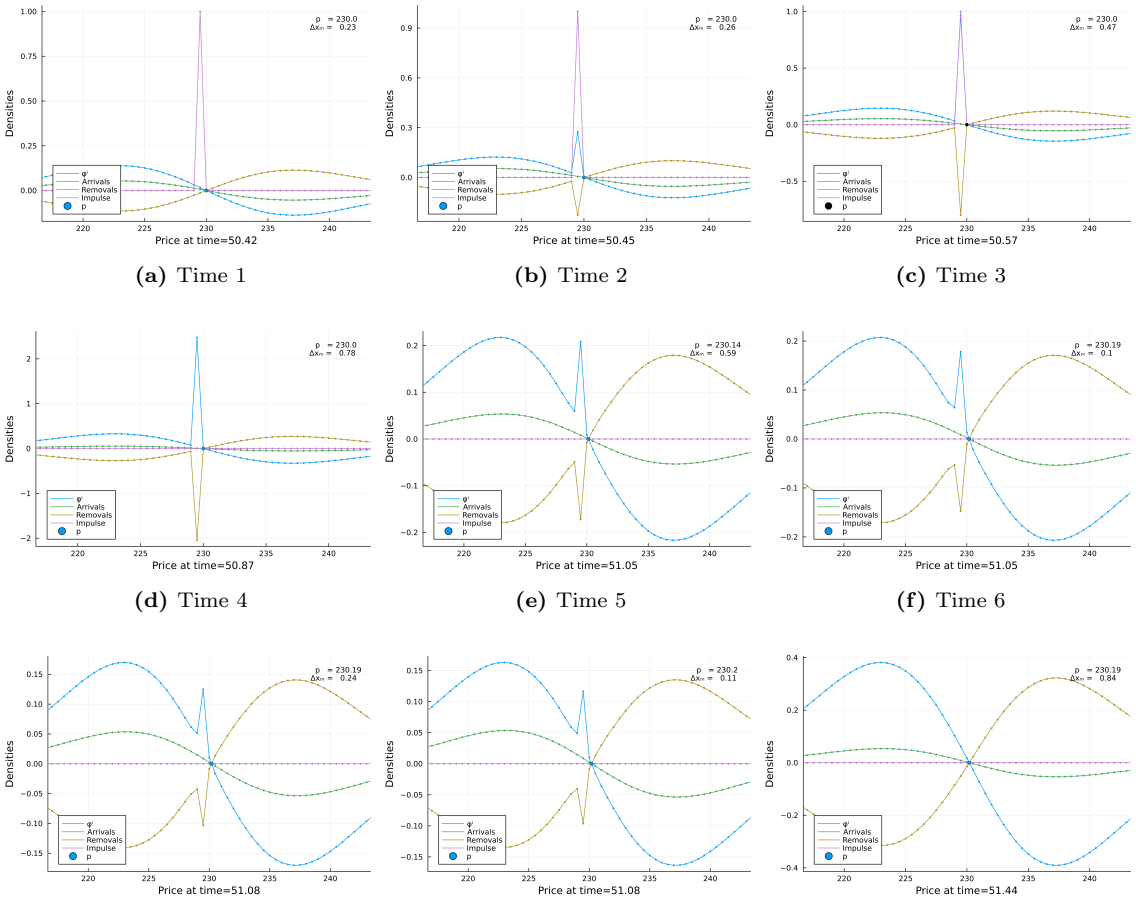


(a) Non-uniform Δt



(b) Uniform Δt

Figure 1: Figures 1a & 1b show the price paths generated using the coupled order book equation in Equation [3.2] with parameters defined in Table 3.2.2 when we couple two order books for the case of non-uniform Δt and uniform Δt respectively (See Section 3.2.1). Refer to Table 4 on how to reproduce these figures.

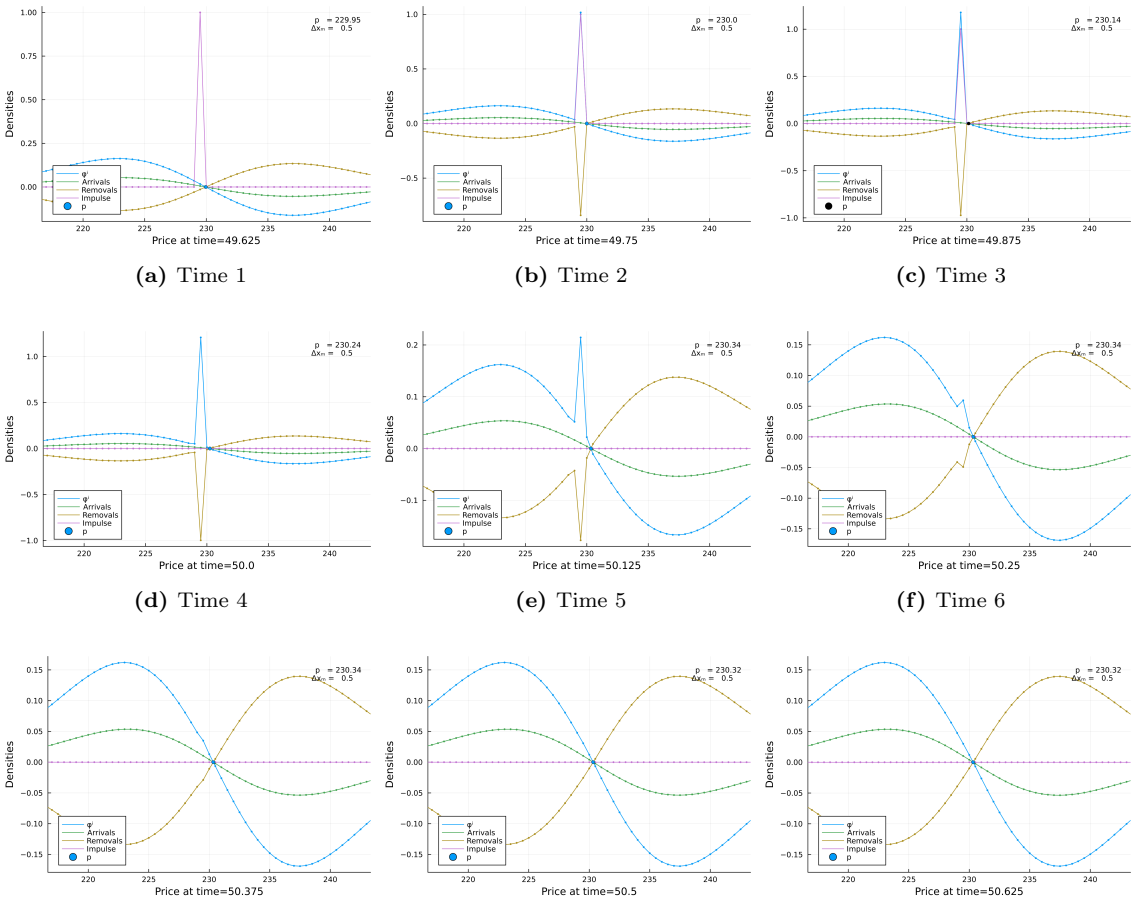


(a) Time 1 **(b)** Time 2 **(c)** Time 3

(d) Time 4 **(e)** Time 5 **(f)** Time 6

(g) Time 7 **(h)** Time 8 **(i)** Time 9

Figure 2: Figures 2a to 2i show a sequence of snapshots depicting the effect of an order book shock on one of the order books which increases price p when we use non-uniform Δt for the coupled order book equation in Equation [3.2] with parameters defined in Table 3.2.2. Since Δt is varied at each time step this results in a different Δx_m for each time step. This corresponds to the price change spike around time $t = 50$ in Figure 1a for Path A. Δx_m and the mid-price p for each snapshot are shown on the top right of each figure. We use pink to indicate an order book shock, gold to indicate order cancellations, green to indicate order arrivals and blue to indicate the mid-price. Note that some of the time increments appear the same but this is as a result of rounding. Where all the lines meet is the equilibrium and forms the mid-price p. Refer to Table 4 on how to reproduce these figures.



(a) Time 1 **(b)** Time 2 **(c)** Time 3

(d) Time 4 **(e)** Time 5 **(f)** Time 6

(g) Time 7 **(h)** Time 8 **(i)** Time 9

Figure 3: Figures 3a to 3i show a sequence of snapshots depicting the effect of an order book shock on one of the order books which increases price p when we use uniform Δt for the coupled order book equation in Equation [3.2] with parameters defined in Table 3.2.2. This corresponds to the price change spike around time $t = 50$ in Figure 1b for Path A. Δx_m and the mid-price p for each snapshot are shown on the top right of each figure. We use pink to indicate an order book shock, gold to indicate order cancellations, green to indicate order arrivals and blue to indicate the mid-price. Where all the lines meet is the equilibrium and forms the mid-price p. Refer to Table 4 on how to reproduce these figures.

4 Epps Effect

Epps (1979) showed that for very short time frames (such as minutes) correlated stocks show very little to no correlation. However, if you observe the correlated stocks over a longer time frame (such as hours or days) they tended toward their expected correlations. In other words, he showed that when you look at stock returns very frequently, they may seem less correlated, but when you aggregate the data over longer time intervals, the correlations become more pronounced. This phenomenon has been corroborated in other studies on equities (Bonanno et al., 2000; Zebedee and Kasch-Haroutounian, 2009; Tumminello et al., 2006; Mastromatteo et al., 2010) and foreign exchange markets (Muthuswamy et al., 2001).

There are three sources which appear to contribute to the Epps effect - asynchrony, lead-lag and tick-size (Chang et al., 2021a), However, these all seem to be artefacts of the fundamentally discrete event nature of real financial markets(Chang et al., 2021b). These sources were explored by others and we highlight some examples here. Münnix et al. (2010) explored the effect that the tick-size has on the Epps effect. Renò (2001) investigated asynchrony when there is a lead-lag present. Precup and Iori (2007) explored varying the levels of asynchrony and showed that this resulted in differently behaving Epps effects. An analytical representation for the Epps effect was shown in Tóth et al. (2007); Tóth and Kertész (2009) and extended by Mastromatteo et al. (2010) to differentiate between the effects of lead-lag and asynchrony.

It is important to realise that Chang et al. (2021b) speculate that the Epps effect should best be understood as a fundamental property arising from the discrete nature of high-frequency financial markets. They questioned whether diffusion processes are an appropriate underlying model representation and that the use of latent models has potentially clouded our understanding of high-frequency finance. They argue that high frequency correlation dynamics can be faithfully recovered when tick data is represented as a web of interconnected discrete events, rather than being samples from continuous Brownian diffusion's as framed by a particular latent model, even when combined with noise.

This has important implications in the current work. Here we show that discrete non-uniform sampling of a continuous-time fluid model generates the Epps effect! This appears to confirm the perspective suggested by Chang et al. (2021b) that discretisation is the primary or foundational cause of the Epps effect, that lead-lags are dressing the underlying impact of discretisation. This can be seen in that an Epps effect arises both in non-uniform and uniform sampling. As long as the system comprises discrete events that do not naturally align in a unique global trading time, when trading in the two order books is discrete, because of discreteness, then there are events that cannot be uniquely aligned when compared cross-sectionally. This means that correlations can only emerge on averaging scales sufficiently long that there are enough events in the discrete sample periods to meaningful allow estimation.

In this section we show that our model, as described in Equation [22], produces this Epps effect phenomenon as shown in Figure 5 for both the uniform and non-uniform Δt cases. We further interrogate the numerical scheme by exploring different different grid point manifestations by changing Δx and Δt in figures 6 & 7 and finally show the effect of having different values for our model parameter ν in Figure 8.

4.1 Correlation estimators

To begin we will first need to estimate the correlations. We implement the non-uniform fast Fourier transform (NUFFT) using the Dirichlet basis kernel with fast

Gaussian gridding (FFG) as described in Chang et al. (2020a). The reasons for using this method are:

i NUFFT methods work well with data that is asynchronous, discrete and event-driven as is the case with our data;

ii the Dirichlet kernel has been shown to plausibly recover the theoretical Epps curve;

iii FFG is computationally efficient.

A pseudo explanation of the algorithm is in Section 7.2 but a high-level overview of it is as follows:

1. Convolve nonuniform source points on the over-sampled grid:

$$f_{\gamma,dp_i}(\xi) = \sum_{h=1}^{n_i-1} \delta_i(I_h)\tilde{\gamma}(\xi_l - t_h^i), \text{for } l = 0, ..., M_r - 1. \tag{33}$$

2. Discrete Fourier Transform (DFT) on the over-sampled grid with standard FFT:

$$F_\gamma(dp_i)(k) = \sum_{l=0}^{M_r-1} f_{\gamma,dp_i}(\xi_l)e^{-2\pi ikl/M_r}, \text{for } k = -\frac{M_r}{2}, ..., \frac{M_r}{2} - 1. \tag{34}$$

3. Deconvolve the effects of the convolution in the Fourier space:

$$F(dp_i)(k) = \frac{F_\gamma(dp_i)(k)}{\tilde{\gamma}(k)}, \text{for } k = -N, ..., N. \tag{35}$$

4.2 Results

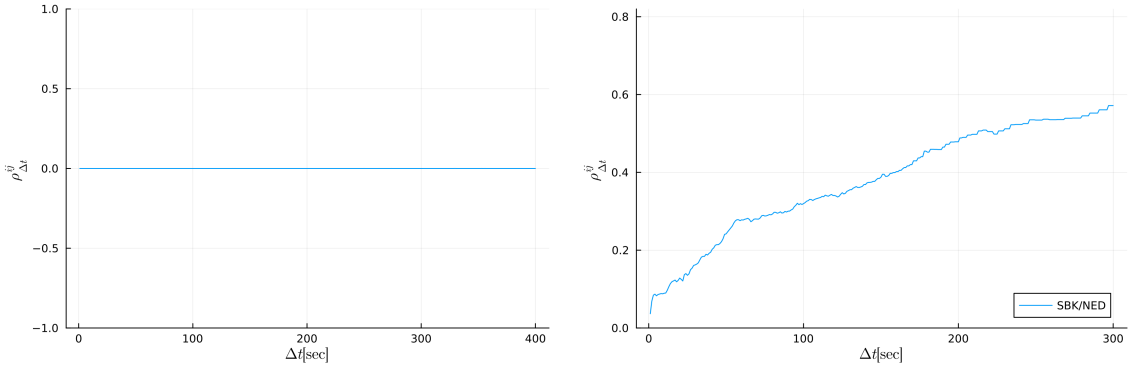
4.2.1 Null and empirical case

In Figure 4a we observe the correlations over varying changes in time scale (Δt) between two coupled Brownian motions. We use this to confirm that if our model does produce an Epps effect it is not due to how we compute our correlations. There is a flat line as Δt increases, confirming that there is no correlation between the two coupled Brownian motions for all Δt. We can thus be confident in any proceeding findings.

In Figure 4b we use empirical data from Chang et al. (2020b) to observe the correlations between two banking stocks listed on the JSE, namely, Standard Bank (SBK) and Nedbank (NED). These are an example of correlated shares which produce an Epps effect (Chang et al., 2021a). We confirm that we are able to reproduce this finding using our method of computing correlations as the correlations form an Epps effect over changes in time scale (Δt).

4.2.2 Initial result

Using the parameters from Table 3.2.2 in our model we are able to produce a price path for each of the two coupled order books for the non-uniform Δt case and uniform Δt case. We then use these price paths to produce Figure 5 which plots the correlations of the price paths (ρ) against changes in time scale (Δt) for non-uniform Δt in Figure 5a and uniform Δt in Figure 5b. We further include the power density curve as an inset in the bottom right of both figures. In Figure 5a we see the emergence of the



(a) Null Case **(b)** Empirical Case

Figure 4: In Figure 4a we use two coupled Brownian motions and estimate their correlations over changes in time scale Δt. We confirm that there is a resulting horizontal line of correlations $\rho \approx 0$, therefore showing that our estimation scheme does not affect the results. In Figure 4b we use daily share data of two correlated bank stocks (Standard Bank - SBK and Nedbank - NED) and show the Epps effect emerges when using the correlation estimation scheme defined in Section 4.1. Our model should produce a similar curve for the correlated order books. Refer to Table 4 on how to reproduce this figure.

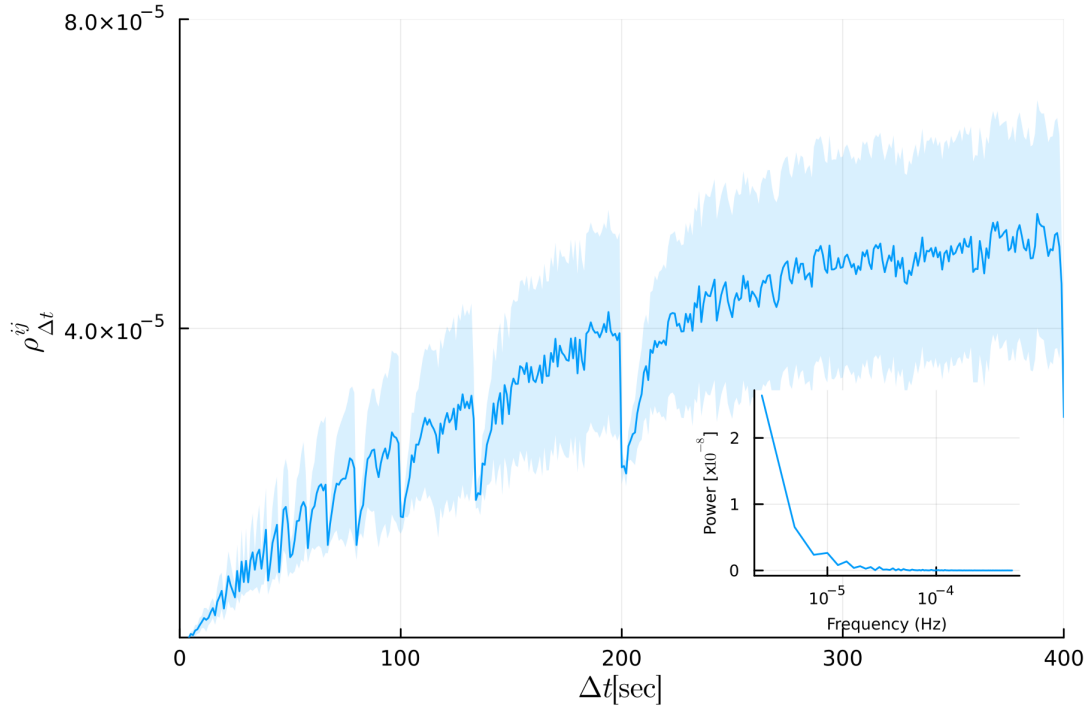
Epps effect, however, there appears to be periodic drops in correlations for different time scales. The power spectrum curve decays with a few minor spikes. Similarly in Figure 5b we see the emergence of the Epps effect however the drops in correlation have a greater magnitude. The power spectrum curve shows more noticeable spikes as it decays. Taking both figures into account we see that these drops in correlation and their magnitude influence the spikes seen in the decay of the power spectrum curve.

4.2.3 Changing Δt

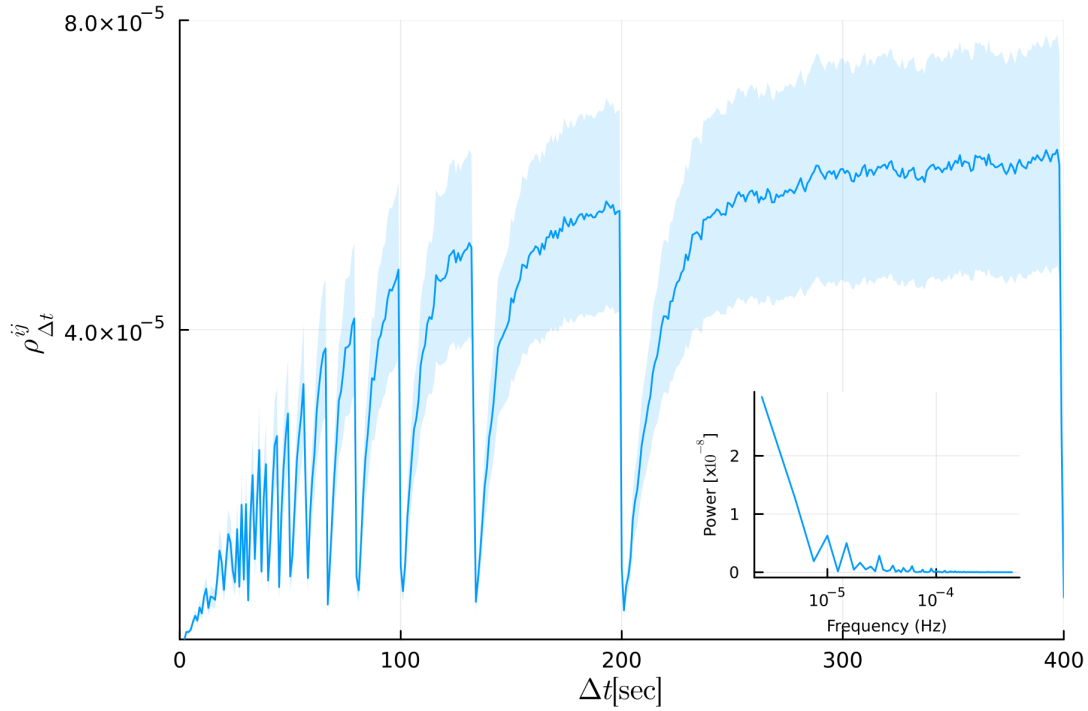
We now explore the effect of changing Δt on the output of the Epps of effect. Figure 6 shows the effect of increasing Δt while keeping a constant $\Delta x = \frac{1}{2}$. We accomplish this by using Equation [32] and changing the value of D, the diffusion rate. We choose to hold r, the probability of a self jump, constant. There are four plots and each has an accompanying power spectrum curve as an inset in the bottom right-hand corner. The first starts with $\Delta t = 0.0625$ ($D = 1.0$) the next has $\Delta t = 0.0781$ ($D = 0.8$) followed by $\Delta t = 0.125$ ($D = 0.5$) and finally we end with $\Delta t = 0.2083$ ($D = 0.3$). The Epps effect is shown quite clearly in all of the plots. However, as with Figure 5, each plot displays instances where there are seemingly random drops in correlation at values of Δt. These drops in correlation appear to increase in magnitude as Δt is increased. The power spectrum curves of each show a spiky decay where the size of the spikes are related to the increase in the magnitude of the drops in correlation.

4.2.4 Changing Δx

Figure 7 shows the effect of changing Δx while keeping a constant $\Delta t = 0.03$ on the emergence of the Epps effect. The are four figures starting with $\Delta x = 0.25$ then $\Delta x = 0.33$ followed by $\Delta x = 0.5$ and finally ending with $\Delta x = 0.75$. Each figure is accompanied by a plot of the power spectrum inset which is at the bottom right-hand corner of the figure. The Epps effect is shown in figures 7a, 7b and 7c with all the figures showing periodic drops in correlation at particular Δt. Figures 7a, 7b appear quite similar compared with Figure 7c which has comparatively much larger drops in

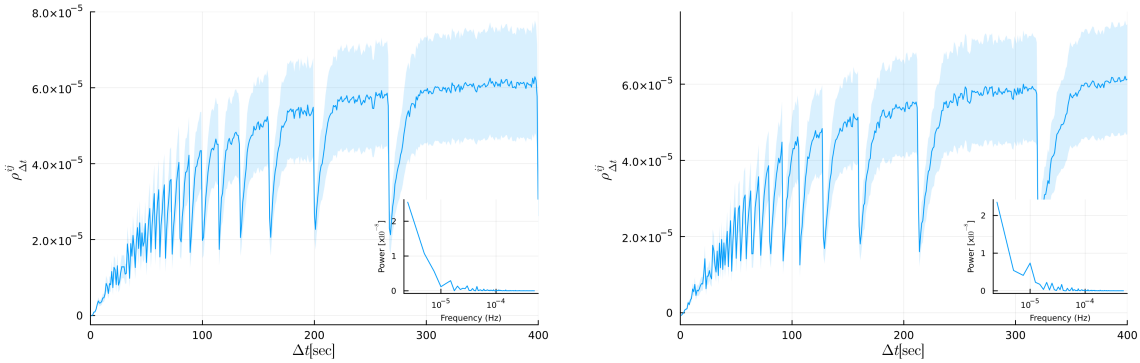


(a) Non-uniform Δt



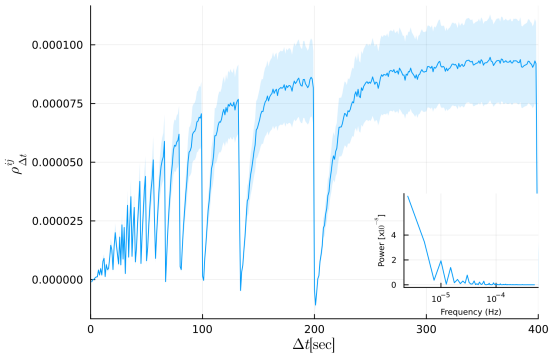
(b) Uniform Δt

Figure 5: For figures 5a, 5b we use our model in Equation [29] with parameters from Table 3.2.2 to generate price paths for two coupled order books. Figure 5a uses non-uniform Δt and Figure 5b uses uniform Δt (See 3.2.1). We then estimate the correlations between the two generated price paths ρ for different time scales Δt by using the NUFFT estimator (See Section 4.1). The correlations have been averaged over 10 iterations. We further include the power spectrum curve which compares power to the log frequency as an inset in the bottom right of the figures. We observe in both that the correlations follow the Epps effect curve with periodic drops in correlation while noting that the magnitude of these drops is a lot bigger in Figure 5b compared with 5a. For the power spectrum curves we observe in both decaying power as frequency increases which indicates there is initially a signal which quickly disappears. We note that 5b has a more noisy decay compared with Figure 5a. Refer to Table 4 on how to reproduce this figure.

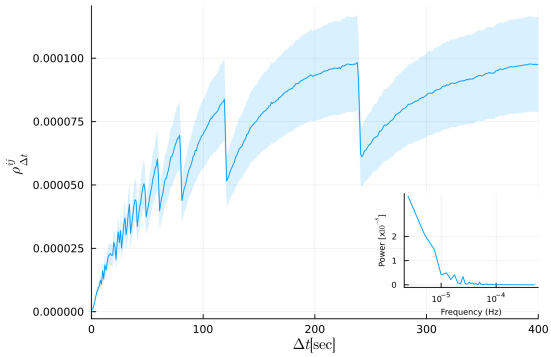


(a) $\Delta t = 0.0625$ $\Delta x = 0.5$

(b) $\Delta t = 0.0781$ $\Delta x = 0.5$



(c) $\Delta t = 0.125$ $\Delta x = 0.5$



(d) $\Delta t = 0.2083$ $\Delta x = 0.5$

Figure 6: We explore the effect changing Δt while keeping $\Delta x = 0.5$ has on the Epps effect. Using Equation [32] we can manipulate the value of Δt by changing D or r or both. We have opted to leave $r = 0.5$ and to change D by setting it to 1 for Figure 6a making $\Delta t = 0.0625$, 0.8 for Figure 6b making $\Delta t = 0.0781$, 0.5 for Figure 6c making $\Delta t = 0.125$ and 0.3 for Figure 6d making $\Delta t = 0.2083$. All other parameters have remained the same as in Table 3.2.2. We have also included the power spectrum curve for each of the figures. We see the Epps effect emerge in each of the figures while each figure shows periodic drops in correlation. Further, the power spectrum curve decays as in Figure 5 for figures 6a, 6d but the decays begin to become increasingly spiky for figures 6b, 6c. These increasing spikes seen in the decay are matched by increases in the magnitude of drops in correlation. Interestingly the magnitude of drops in correlation increase as Δt increases in figures 6a,6b and 6c but then is reduced in Figure 6d. Each of the four figures shows correlations which have been averaged over 10 iterations. Refer to Table 4 on how to reproduce these figures.

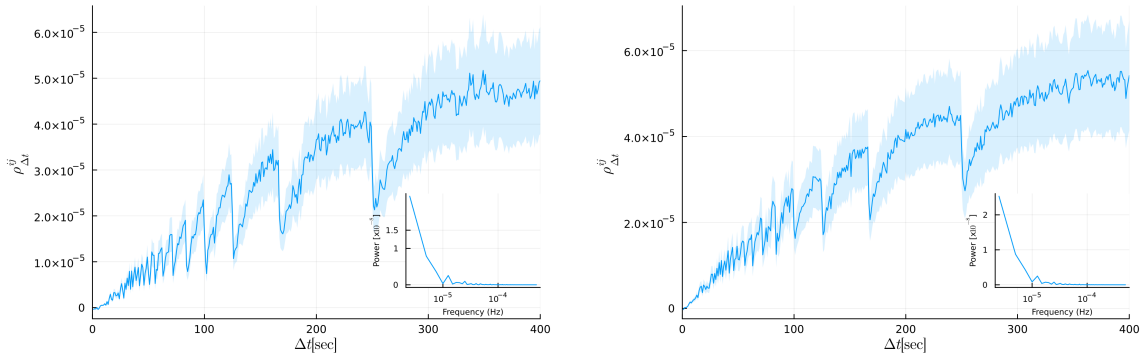
correlation. Again we see the power spectrum in each decay with spikes which are increased in magnitude based on the increased drops in correlation. Figure 7d does not show the Epps effect and in fact, shows a decreasing trend in correlation ρ for changes in time scale Δt.

4.2.5 Changing ν

We try to explain the drops in correlation seen previously by exploring the effects of changing ν.

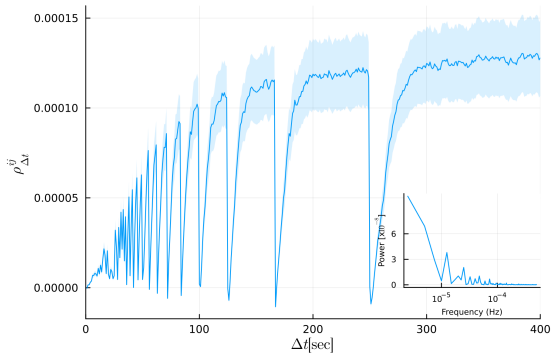
Figure 8a examines the effect of changing ν on the Epps effect when we use non-uniform Δt combined with the parameters from Table 3.2.2. We show the Epps effect for values of ν starting at $\nu = 1$ and incrementing by 3 until we reach a value of $\nu = 16$. We see that at $\nu = 1$ we are able to recover the Epps effect without any drops in correlation. As ν increases we start to see the drops in correlation previously seen and there appears to be a positive correlation between the magnitude of ν and the magnitude of the drop in correlation. In other words, as ν is increased the drops in correlation increase in size.

Figure 8b examines the effect of changing ν on the Epps effect when we use uniform Δt combined with the parameters from Table 3.2.2. We show the Epps effect for values of ν starting at $\nu = 1$ and incrementing by 3 until we reach a value of $\nu = 16$. Similarly to the non-uniform Δt case, we see that at $\nu = 1$ we are able to recover the Epps effect without any drops in correlation. As ν increases we start to see the drops in correlation previously seen. These drops in correlation have a far greater size than in the case of non-uniform Δt. Again, there appears to be a positive correlation between the magnitude of ν and the magnitude of the drop in correlation. In other words, as ν is increased the drops in correlation increase in size.

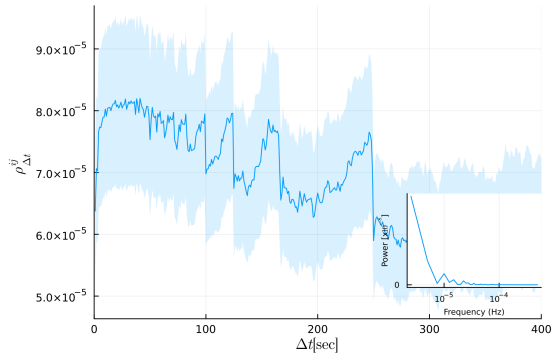


(a) $\Delta t = 0.1$ $\Delta x = 0.25$

(b) $\Delta t = 0.1$ $\Delta x = 0.33$

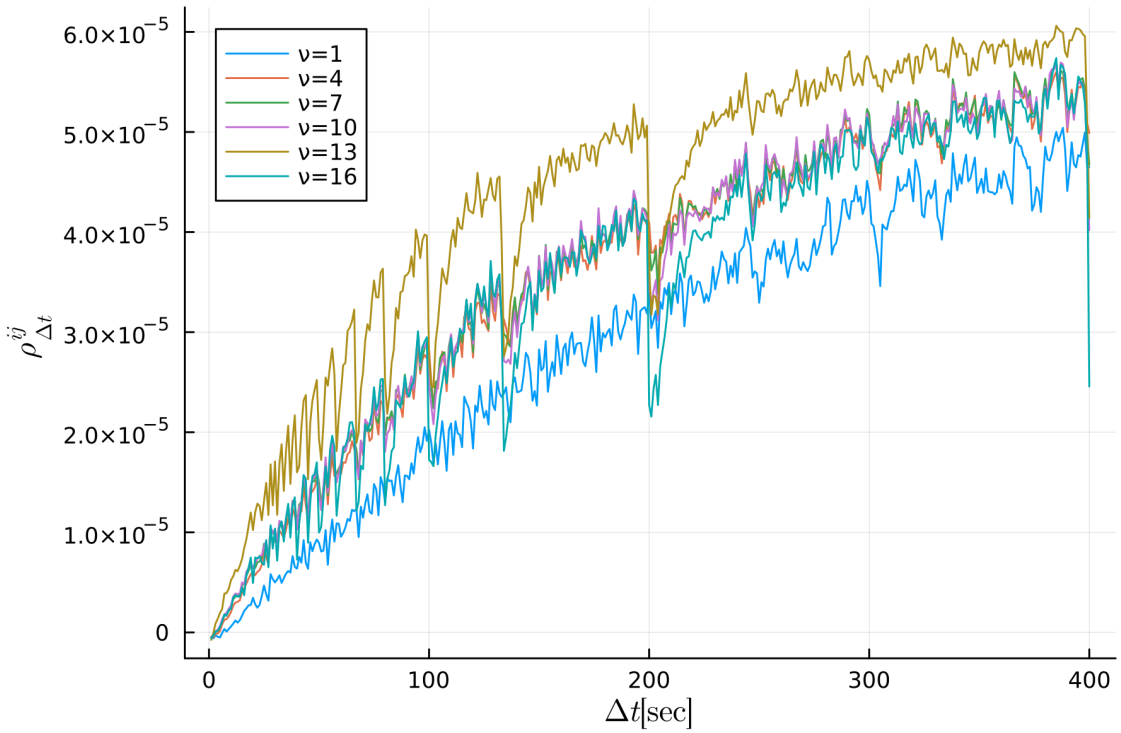


(c) $\Delta t = 0.1$ $\Delta x = 0.5$

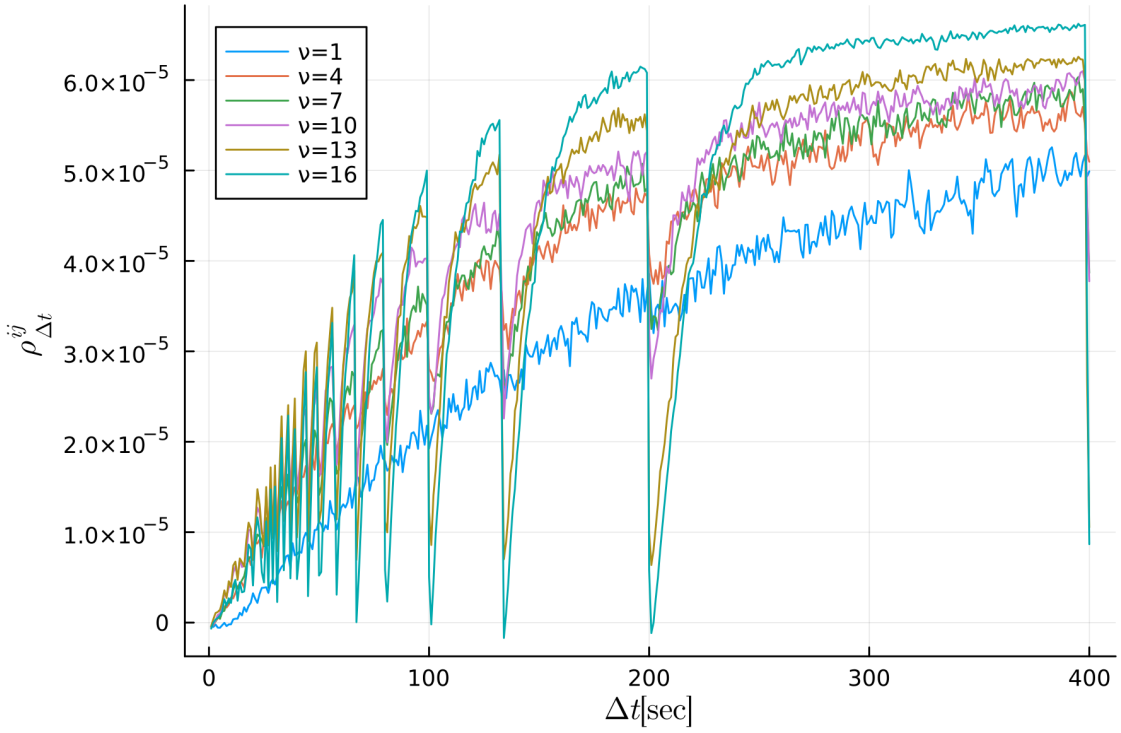


(d) $\Delta t = 0.1$ $\Delta x = 0.75$

Figure 7: We explore the effect of changing Δx while keeping $\Delta t = 0.1$ on the Epps effect. Given that $\Delta x = \frac{L}{M}$ we can manipulate its value by changing L or M or both. For Figure 7a we have decreased $L = 100$ so that $\Delta x = 0.25$, for Figure 7b we have decreased $L = 100$ and $M = 300$ so that $\Delta x = 0.33$, for Figure 7c we have kept the values the same and for Figure 7d we have increased $L = 300$ so that $\Delta x = 0.75$. We have also included the power spectrum curve for each of the figures. Similar to Figure 6 we see the Epps effect emerge for figures 7a, 7b and 7c with each showing periodic drops in correlation. However for Figure 7d there is no Epps effect with the curve decreasing over increases in time differences. Further, and again, the power spectrum curve initially decays as in Figure 5 for figures 7a, 7b but the decay becomes increasingly spiky for Figure 7c. These increasing spikes seen in the decay are matched by increases in the magnitude of drops in correlation. Interestingly the magnitude of drops in correlation increase as Δx increases from figure 7b to 7c. Each of the four figures shows correlations which have been averaged over 10 iterations. Refer to Table 4 on how to reproduce these figures.



(a) Non-uniform Δt



(b) Uniform Δt

Figure 8: We further investigate the reason for the drops in correlation by comparing the Epps effect when we use different values of ν when we use uniform and non-uniform Δt (see 3.2.1) for figures 8a, 8b respectively. We start at $\nu = 1$ and increase in increments of 3 until $\nu = 16$. The rest of the parameters remain the same as those from Table 3.2.2. We show that the magnitude of the drop in correlations appears to be related to an increase in ν. Each of the correlation curves using different values of ν has been averaged over 10 iterations. Refer to Table 4 on how to reproduce these figures.

5 Calibration

We now seek to improve our model parameters by trying to calibrate them using empirical data. Given the model is simulating intraday trading, we have opted to follow the insights found in the agent-based modelling (ABM) literature. In the early stages of ABM research, parameters for the model were chosen manually to highlight they could generate very specific statistical characteristics in the simulated data *e.g.* log-return stylised facts. However, there was no assessment to determine if the distribution of the generated data matched that of real empirical data, and no demonstrations of the extent to which parameter choices were pathologically chosen to unreasonably tune the model; where they would otherwise be unstable under small parameter changes (have large parameter sensitivity and degeneracies), or merely reflecting instability in the choice of objective functions in the presence of multiple similar local optima (objective function degeneracy). This implied that many of the early attempts in the ABM literature to claim this or that insight were not statistically meaningful, and often did not always faithfully reflect the models themselves. This created a calibration challenge, where the model needed to be fine-tuned to better match real-world data (Fabretti, 2013; Platt and Gebbie, 2017) in a more statistically believable manner (Platt, 2020).

To address this calibration problem, researchers typically employ one of three common methods for financial ABMs (Platt and Gebbie, 2017).

1. Maximum Likelihood Estimation: This method is suitable for models with closed-form solutions, particularly for daily sampled data and approximating closing auctions. However, it is not appropriate for models involving intraday trading.
2. Bayesian Inference: Bayesian inference constructs a posterior distribution of the model's parameters using techniques like kernel density estimation or parametric approaches. While this method can outperform other techniques in various scenarios (Platt, 2020), it is computationally intensive.
3. Method-of-Moment using Simulated Minimum Distance (MM-SMD): This method is chosen in the context described because it is computationally manageable, and does not require closed-form solutions. MM-SMD involves matching moments of the simulated data to those of empirical data, making it a practical choice for calibration.

It is important to note that MM-SMD has its limitations (Platt, 2020). Critics have pointed out that it can lead to parameter degeneracies and unpredictable impacts on the optimisation process, causing it to converge to local optima (Platt and Gebbie, 2017). Furthermore, MM-SMD relies on the selection of specific moments that may only represent a limited set of aggregate properties in the data, potentially missing important dynamics. The method can also have a "degrees-of-freedom problem" when there are not enough moments relative to parameters and the intra-parameter relationships and degeneracies.

Despite these shortcomings (as an estimation method), MM-SMD remains a quick, straightforward, and reliable method for calibrating intraday trading models. Particularly when there are more moments than parameters. This is why it is our chosen method of calibration. However, it is crucial to consider the calibrated parameters as indicative rather than robust (that is useful for inference) due to these inherent issues (Jericevich et al., 2021a).

5.1 Data

When using the empirical data we needed to filter and prepare the data for analysis. This involved discarding data related to the opening and closing auctions, focusing only on trading activity that occurred between 9:00 and 16:50. Additionally, we observed that the first minute of trading produced unreliable data, so we decided to exclude it from our analysis to facilitate the model calibration.

To ensure the data was clean and relevant, we removed events associated with intraday volatility auctions and the impact of various futures close-outs. Since futures close-outs only occur in March, June, September, and December, we didn't need to consider these cases. We also eliminated unwanted trades, such as after-hour trades (LT), corrections of the previous day's trades (LC), and auction uncrossing price trades (IP), focusing solely on automated trades (AT).

In financial markets, it's common for larger trades to be executed as a combination of smaller trades. This can lead to a single trade at a specific timestamp being split into multiple smaller trades, impacting the best bid and best ask prices as they fluctuate in the order book. The data may represent this event as multiple trades when it's actually a single trade. Additionally, some timestamps may have multiple quotes associated with them.

To address these complexities in the data, we conducted trade and quote compacting. This process involved modifying the data to better reflect the occurrence of a single trade and ensuring that there was only one quote per timestamp. When multiple quotes shared the same timestamp, we retained the most recent quote and removed the others. For trades with the same timestamp and the same order type, we calculated the aggregated trade volume for that timestamp and determined the price as the volume-weighted average price. This step was essential for cleaning and simplifying the data for our analysis.

5.2 Selection of moments

In the process of constructing MM-SMD, we must first select our moments. We use the recommendation put forth by Winker et al. (2007) to use a substantial number of moments, encompassing a wide array of characteristics drawn from empirical data. This selection serves to improve the chances of accurately identifying the parameters of our model. The moments chosen for this purpose are shown in Table 2. These moments have been selected due to their ability to encompass stylised facts that we aim to reproduce within our model.

Notably, we deviate from the prior study in Jericevich et al. (2021a) by excluding kurtosis as one of the chosen moments. Empirical research has shown the unreliability of kurtosis as a moment (Winker et al., 2007). Jericevich et al. (2021a) explained the limited reliability of kurtosis, citing instances where it generated elevated values attributable to extreme initial price changes, subsequently leading to prolonged price stagnation.

To evaluate the fat-tailed characteristics of the return distribution, we employ the "improved" Hill estimator to determine the tail-index of the upper 5% of the right tail of the log-return distribution (Winker et al., 2007). The standard deviation of log-returns is retained as a moment, given its recognition as a robust statistical measure with a higher likelihood of presence compared to kurtosis.

Furthermore, we base the estimation of these moments on micro-price log returns, as opposed to mid-prices. We decided on this, as Jericevich et al. (2021a) have empirically demonstrated that employing mid-prices for this purpose leads to unreliable outcomes in the characterization of distributional properties.

Moments	Targeted Empirical Feature	Description
Mean, Std.	Distribution shape	The mean, μ, and standard deviation, σ, measure the shape of the log-return distribution.
KS	Distribution shape	The Kolmogorov-Smirnov (Massey, 1951) (KS) statistic is used to compare the maximum absolute difference between the CDFs of the empirical and simulated log-returns. This is used to test if the samples are drawn from the same distribution.
GPH	Long-range dependence	Geweke and Porter-Hudak (1983) (GPH) provide an estimate for the fractional integration parameter $d \in [-0.5, 0.5]$ in absolute log-returns time series by estimating the memory parameter in the log-periodogram. This measures the long-range dependence in log-returns.
ADF	Serial auto-correlation	The Augmented Dickey-Fuller (Dickey and Fuller, 1979) (ADF) statistic tests for the presence of a unit root in the time series of the log-returns. The more negative the test statistic the less evidence there is for the presence of a unit root.
GARCH $\alpha + \beta$	Short-range dependence	The parameters of the GARCH(1,1) model provide a useful summary for the short-range dependence in the squared returns. Winker et al. (2007) found that the sum of the parameters in the GARCH(1,1) model is a more robust statistic as compared to the individual parameters.
HE	Power-law tail	The Hill Estimator (HE) is used to estimate the tail-index of a power-law distribution which can be used to infer the power-law behaviour in the tails of experimental distributions. Nuyts (2010) proposed an "improved" Hill estimator to remedy the issues with the standard estimator.
H	Long-memory	The Hurst (H) exponent (Hurst, 1951; Mandelbrot and Wallis, 1969) provides a measure of the long-memory of the log-returns. If $H \in [0, 1]$ is significantly far from 0.5 we have evidence that the log-returns do not follow a random walk. Furthermore, if $H \in [0, 0.5)$ then the process is said to be mean-reverting and if $H \in (0.5, 1]$ the process is said to be trending.

Table 2: A list of all the moments used for calibration with the features they calibrate for and a description of the moment adapted from Jericevich et al. (2021a); Dicks et al. (2023).

5.3 Simulated minimum distance

The calibration problem involves optimising a stochastic approximation of an objective function where we aim to minimise the function f($\boldsymbol{\theta}$) - $\boldsymbol{\theta}$ represents a parameter vector within the parameter space θ, and f stands for the objective function itself. In our case, the parameters we wish to optimise are $\boldsymbol{\theta} = [D, \nu, \gamma]$ which are detailed in Table 3.2.2 and are our free parameters. The other parameters L and M were initially also free however in practice allowing them to change added negative or no value to improving the Epps effect or the the stylised facts and thus were kept constant.

The MM-SMD method takes into account three categories of moments: the true moments $\boldsymbol{m} = (m_1, ..., m_d)$, the estimated moments from empirical intraday price data $\boldsymbol{m^e} = (m_1^e, ..., m_d^e)$, and the moments estimated from simulated price paths given specific parameters $\boldsymbol{\theta}$, $\boldsymbol{m^s|\theta} = (m_1^s, ..., m_d^s)$. The true moments $\boldsymbol{m}$ are the actual moments of the selected moments of log-returns or features, which cannot be known directly. The estimated moments $\boldsymbol{m^e}$ are derived from the empirical intraday price path data. The moments $\boldsymbol{m^s|\theta}$ can be estimated from the simulated price path of the model given a particular set of parameters θ. Using the standard method of moments condition $\mathbb{E}[\boldsymbol{m^s|\theta}] = \boldsymbol{m}$. If we know $\boldsymbol{m}$ then $\mathbb{E}[\boldsymbol{m^s|\theta - m}] = \boldsymbol{0}$ and given I number of Monte Carlo simulation iterations we have that $\frac{1}{I}\sum_{i=1}^{I}[\boldsymbol{m_i^s|\theta - m}] = 0$. As we do not know the true moments $\boldsymbol{m}$, we use the estimated moments $\boldsymbol{m^e}$ instead. We define an error measure E in terms of the distance between the simulated moments and the estimated moment:

$$\hat{E}(\boldsymbol{\theta}) = \frac{1}{I}\sum_{i=1}^{I}[\boldsymbol{m_i^s|\theta - m^e}] \tag{36}$$

The number I of Monte Carlo replications are key to reducing the variance of the stochastic approximation of the objective function. We set $I = 5$ as it is more computationally practical, however, it should be noted that a number greater than 5 is ideal to reduce the effects of randomness.

We are able to construct the full optimisation function as a Weighted Sum of Square Errors in E with a matrix of weights $\boldsymbol{W}$ as follows:

$$f(\boldsymbol{\theta}) = \hat{E}(\boldsymbol{\theta})^T \boldsymbol{W} \hat{E}(\boldsymbol{\theta}) \tag{37}$$

Following Winker et al. (2007), we set $\boldsymbol{W} = \mathbb{C}\text{ov}[\boldsymbol{m^e}]^{-1}$, as this intuitively applies larger weights to moments with greater uncertainty (Platt and Gebbie, 2017). We estimate $\boldsymbol{W}$ using a moving block bootstrap, as this is meant to preserve the auto-correlation in the time series. The window size used is 2000.

5.4 Optimisation

Due to computational constraints, we have opted to use the Nelder and Mead (1965) (NM) optimisation algorithm. In the context of optimising a set of n-free parameters, the NM algorithm constructs a simplex consisting of $n + 1$ vertices. Each vertex represents a potential solution within the optimisation space. During each iteration, the simplex undergoes a transformation of either a reflection, expansion, contraction, or shrinkage operation. These transformations are carried out repeatedly until one of the following conditions is met: either the convergence criteria are satisfied, or the optimisation process is terminated due to a stopping condition.

To avoid a rapid descent into local minima resulting from the stochastic approximation of the objective function we have combined the Threshold Acceptance (TA)

algorithm with the NM algorithm. The TA algorithm involves a process where it selects a single point within the objective function and introduces perturbations to create a new point. If this new point yields a superior objective function value or is within a threshold τ of the current solution, it is adopted as the new solution (Winker et al., 2007). This inclusion of somewhat suboptimal solutions allows the algorithm to break free from local optima, enhancing its exploratory capabilities. It is worth noting that the threshold value τ is progressively reduced during the optimisation process, facilitating convergence. Each value of τ is employed for a predetermined number of steps, ensuring that the algorithm traverses the parameter space under various threshold conditions multiple times, as explained in the work by Jericevich et al. (2021a).

The combination of the Threshold Accepting (TA) and Nelder-Mead (NM) algorithms results in the NMTA algorithm. In the NMTA algorithm, at each iteration, a decision is made to execute either a single step from the TA algorithm or a single step from the NM algorithm. The choice between a TA or NM algorithm step is probabilistic, with TA being selected with a probability denoted as ε and NM with a probability of 1 - ε. NM is favoured with a higher probability. The TA algorithm is applied to the vertex with the lowest objective function value within the simplex. The magnitude of perturbation applied in the TA algorithm is contingent on the mean value calculated over the entire simplex for the parameter under consideration, as detailed in the study by Jericevich et al. (2021a). Furthermore, it is important to note that during the operations in the simplex search, the objective function values are also subjected to comparison using the same thresholding criteria as in the TA algorithm. This ensures that suboptimal solutions are appropriately considered and selected within the NM algorithm.

During initial testing using the NMTA algorithm, we found that the values required bounds to observe better results in terms of the stylised facts and the Epps effect. The bounds used are as follows: $\nu \in [0, \infty)$, $D \in [0, \infty)$, $\gamma \in [0.4, 1]$. We ran 100 iterations of the algorithm on each attempt at calibration. The values were found to be highly sensitive and convergence changed significantly during each attempted run of the NMTA algorithm. As such the final values can not conclusively be considered as the most optimal values.

5.5 Results

Table 3.2.2 shows the calibrated values found using the NMTA algorithm with associated confidence intervals. The magnitude in variation of the confidence intervals for D and ν show that can have very little confidence in values obtained. In contrast the confidence interval for γ has a lot less variation and as such we have more confidence in its value.

Using the values from Table 3.2.2 we generate the figures in Figure 9. Both figures show the correlations over different time scales with the power spectrum as an inset in the bottom right hand corner of the figure. For Figure 9a we use non-uniform Δt. We see clearly the emergence of the Epps effect and notably without any drops in correlations such as the figures seen in Section 4. Further the power spectrum decays without any spikes providing more of an indication of the relationship between the spikes in the power spectrum with the drops in correlation. Given the calibrated value of $\nu = 12.55$ and based on the results of Figure 9a we would have expected there to definitely be drops in correlation. However, it appears the decrease in γ combined with the decrease D resulting in the increase in Δt have curbed this.

Figure 9b shows the Epps effect when we use the model with uniform Δt. Again

				Calibrated Values		
Parameter	Description	Type	Range	$\Theta_{0.025}$	Θ	$\Theta_{0.975}$
L	System length	Fixed			200	
M	Number of divisions	Fixed			400	
r	Probability of self jump	Fixed			0.5	
D	Diffusion constant	Free	$[0, \infty)$	-4.018	0.27	4.558
ν	Cancellations rate	Free	$[0, \infty)$	-21.635	12.55	46.735
γ	Fraction of derivative	Free	$[0.4, 1]$	0.484	0.57	0.656
$p_{1,2}(0)$	Initial prices	Fixed			230	
$\lambda_{1,2}$	Source terms intensities	Fixed			1	
$\mu_{1,2}$	Source terms rates	Fixed			0.1	
Δx	Change in x ($\frac{L}{M}$)	Fixed			0.5	
Δt	Change in t (see Equation [32])	Free			0.2315	

Table 3: The calibrated model parameters with description and bounds. When running the NMTA algorithm we found the calibrated values for D, ν and γ along with their confidence intervals. Both D and ν have large confidence intervals which indicate a very high variability and as such less trust in the actual value, while γ has less variability indicating more trust in the actual value. The value for Δt has changed compared to Table 3.2.2 as a result of the change in D. We use these parameters for the Epps effect figures 9a & 9b and the stylised effects in Figure 10.

we see the periodic drops in correlation with the associacted spikey decay in the power spectrum. Interestingly we do not see the decrease in γ combined with the decrease D resulting in the increase in Δt having any effect on the drops in correlation as we did with the non-uniform Δt case.

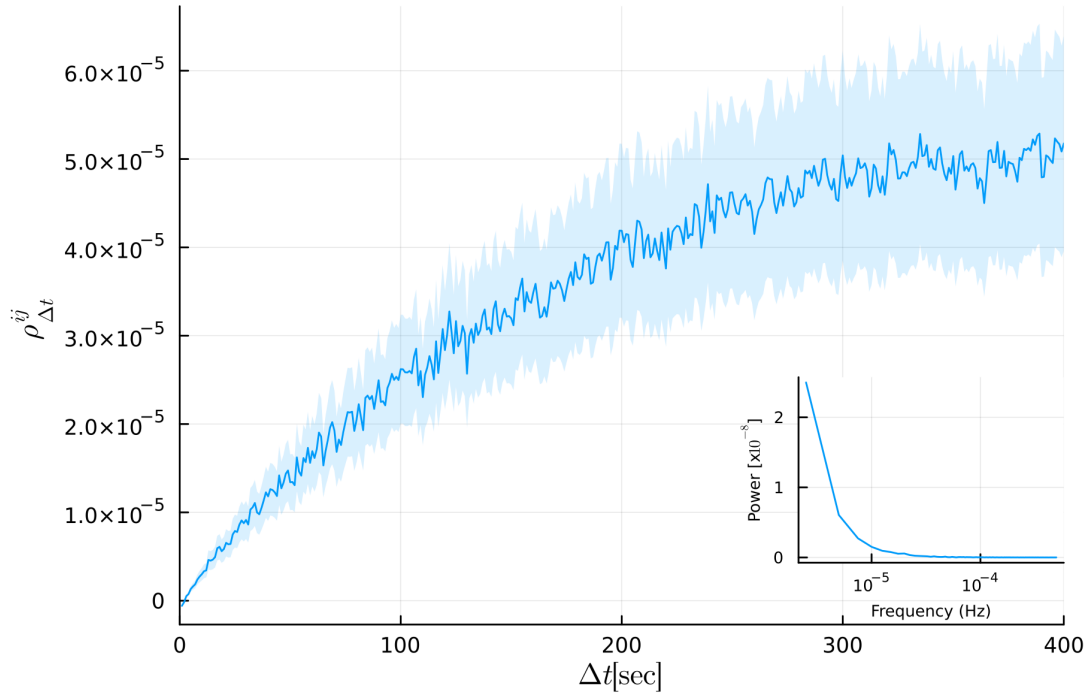
The stylised facts for the empirical data along with both the non-uniform and uniform Δt cases are shown in Figure 10. Each of the cases has three accompanying stylised facts figures. Each of the three cases (i.e. empirical, non-uniform and uniform Δt) going from the left most figure to right most has a price curve with a log-returns inset in top right hand corner, followed by the log returns distribution with a QQ-plot of the returns in the top right hand corner and a figure with three autocorrelation curves (ACF) for the log returns, absolute log returns and order flow. Figures 10a, 10b and 10c show the stylised facts for the empirical results. The output of these figures is what we would like our model to reproduce. In other words we want a realistic looking price curve, noisy log-returns, log-returns which follow a normal distribution, an S-like QQ plot shape with decaying autocorrelations for absolute log returns and order flow.

Figures 10d, 10e and 10f are the stylised facts for the non-uniform Δt case. We see a realistic price curve with noisy log returns in Figure 10d. We see a normal distribution of log returns and an almost S-like QQ plot in Figure 10e. We only see an initial autocorrelation decay for the ACF of absolute log returns and no decay for the order flow in Figure 10f.

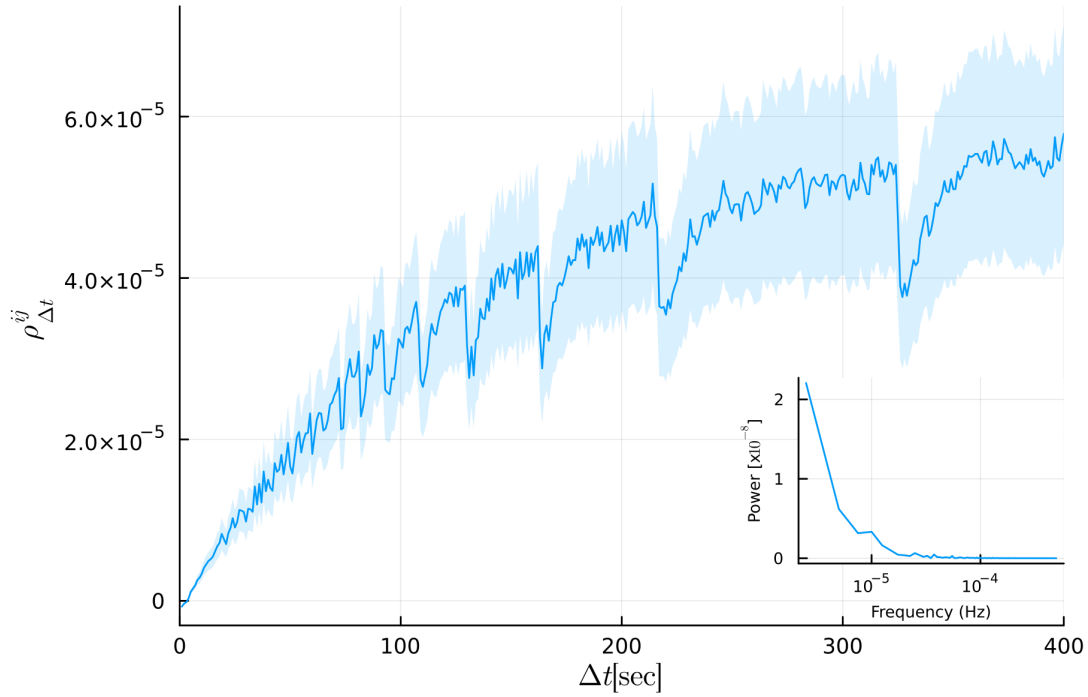
Figures 10g, 10h and 10i are the stylised facts for the uniform Δt case. We see a realistic price curve with noisy log returns in Figure 10g. It is worth noting that we have an outlier in the log returns early on, where we see a straight vertical line. We see a normal distribution of log returns and an almost S-like QQ plot in Figure 10h with the previously mentioned notable outlier for the log returns extending the positive side of the distribution and showing up in the top right of QQ plot inset. We see a very minor initial autocorrelation for the ACF of absolute log returns and no

autocorrelation for the order flow in Figure 10i.

Based on the output of the Epps effect without any drops in correlation accompanied by most of the stylised matching the empirical stylised facts it is clear that the non-uniform Δt case is superior to the uniform Δt case. However, the fact that the autocorrelations do not match up well to the empirical autocorrelations is something that most likely can be improved in future work.

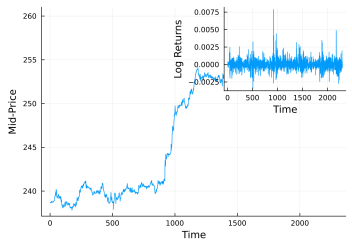


(a) Non-uniform Δt

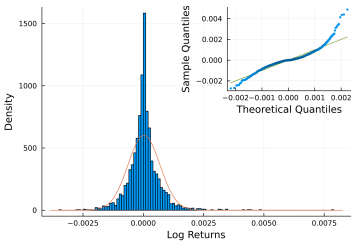


(b) Uniform Δt

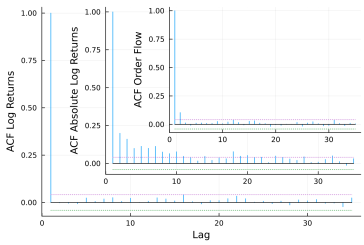
Figure 9: Using the calibrated parameter values found in Table 5.5 we generate the Epps effect for non-uniform Δt in Figure 9a and uniform Δt in Figure 9b with an accompanying power spectrum inset. In Figure 9a we see quite a smooth Epps effect with a similarly smooth decaying power spectrum. This is unexpected given that ν has a large value of 12.55. In Figure 9b we see the periodic drops in correlation as noted earlier and a more choppy decaying power spectrum. This indicates that using the non-uniform Δt case is optimal to produce the better Epps effect. The correlations of each figure have been averaged over 10 iterations. Refer to Table 4 on how to reproduce these figures.



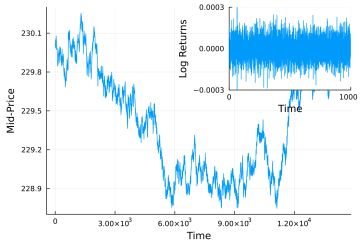
(a) Empirical Price Path & Returns



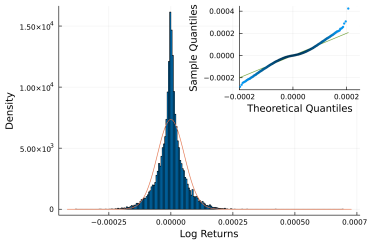
(b) Empirical Returns Distribution & QQ plot



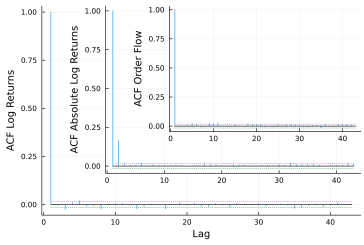
(c) Empirical Autocorrelations



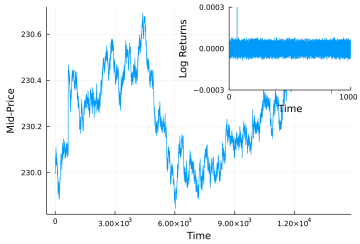
(d) Non-uniform Δt Price Path & Returns



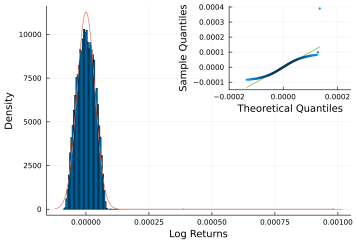
(e) Non-uniform Δt Returns Distribution & QQ plot



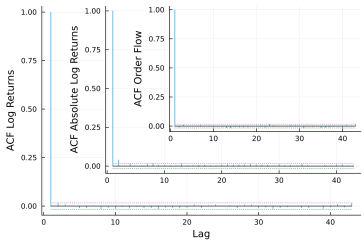
(f) Non-uniform Δt Autocorrelations



(g) Uniform Δt Price Path & Returns



(h) Uniform Δt Returns Distribution & QQ plot



(i) Uniform Δt Autocorrelations

Figure 10: The stylised facts from the empirical data (figures 10a, 10b & 10c), and the data generated when using non-uniform Δt (figures 10d, 10e & 10f) and uniform Δt (figures 10g, 10h & 10i). If our model is accurate it should produce a similar output of stylised facts as those shown in figures 10a, 10b & 10c. In terms of the price returns, both Figure 10d and Figure 10g produce noisy returns and realistic price curves. Figure 10e and Figure 10h show a normal returns distribution which matches the expected output, however, the non-uniform Δt case produces a closer match. We expect to see auto correlations decay and both Figure 10f and Figure 10i show an initial decay but not for as long as seen in the empirical case in Figure 10c. Refer to Table 4 on how to reproduce these figures.

6 Conclusion

This paper aimed to evaluate if the model proposed by Diana and Gebbie (2023a) produced the Epps effect phenomena, which is an empirical characteristic found in correlated financial assets, when the correlations over increasing time scales was measured.

In Section 2 we showed the coupled order book equation given in Equation [22] can be discretised and numerically solved using an update equation (see Section 2.1) with zero-flux boundary conditions (see Section 2.3) and initial conditions (see Section 2.5) to produce the coupled order book equation (see Equation [29]).

In Section 3 we introduced the coupled order book model defined in Equation [22] with the coupling term defined in Section 3.1.1. We noted that we can implement non-uniform Δt and a uniform Δt case for our model. Figure 1 is an example price paths generated when using Equation [29] applied to a two-coupled order book scenario for both cases. We further showed time snapshots of what occurs when there is an order book shock for both the non-uniform Δt case in Figure 2 and uniform Δt case in Figure 3.

In Section 4 we explain the Epps effect and implement the non-uniform fast Fourier transform using the Dirichlet basis kernel with fast Gaussian gridding (see Section 4.1) to estimate correlations between the price paths for different time scales. We verify a null case to confirm that our estimation scheme does not obscure our results and show the Epps effect emerge for empirical data (see Section 4.2 and Figure 4). Thereafter we show that the coupled price paths do produce the Epps effect for both the non-uniform Δt and uniform Δt cases as shown in Figure 5, however, there are periodic drops in correlation found in both. We tried to understand the cause of these drops in correlation which resulted in us evaluating the effects of changing Δt in Figure 6 and Δx in Figure 7 on the Epps effect while holding the other parameters constant. We found that increasing Δt lead to an increase in the magnitude of the drops in correlation and found a similar result for an increase in Δx, however we seem to lose the Epps effect when Δx is too large as shown in Figure 7d. We then tried to vary ν for the non-uniform Δt and uniform Δt cases as shown in Figure 8 and we had success showing that increasing ν resulted in increases in the size of the period drops in correlation, with the uniform Δt case showing larger decreases compared with the non-uniform Δt case. We found that a value of ν close to 1 given the other parameters are held constant results in no drops in correlation for both cases.

Finally in Section 5 we use empirical data (see Section 5.1) and the method of moments (see Section 5.3) combined with the Nelder-Mead Threshold Acceptance optimisation algorithm (see Section 5.4) to calibrate the model parameters, the output of which is shown in Table 5.5. We found that the calibrated values for $D = 0.27$ and $\nu = 12.55$ and a large confidence interval indicating that their values are not very reliable, whereas the value of $\gamma = 0.57$ was found to be within a reasonable confidence interval and therefore a more reliable value. We found that this combination of parameters produced a desired Epps effect for the non-uniform Δt case in Figure 9a without any drops in correlation which came as a surprise given the large value of ν. The uniform Δt case produced a similar plot in Figure 9b to the previous ones with an Epps effect that had periodic drops in correlation. We further compared the stylised facts produced by both cases to the empirical stylised facts we wish to reproduce in Figure 10. The non-uniform case performed best as it resulted in stylised facts which more closely resembled the empirical effects, however, improvements can still be made with regards to the autocorrelations.

In conclusion, we showed that the model proposed by Diana and Gebbie (2023a)

does indeed produce the Epps effect when using the calibrated model parameters with non-uniform Δt. However improvements can be made in future iterations to identify the exact cause of the periodic drops in correlation as this is still inconclusive and for the stylised facts to more closely match the autocorrelations found in empirical data.

7 Algorithms

7.1 Stochastic finite difference algorithm

Algorithm 1 Diffusion simulation algorithm

Require: $\{M, x_0, x_M, r, \lambda_t, D_j, \nu_j, \beta, \lambda_j, \mu_j\}$ (Table 3.2.2)

1: Compute Initial Conditions: φ_0^i
2: Compute Time Steps: $\{\Delta t_k\}_{k=1}^M$ and Δt
3: Compute State Steps: $\{\Delta x_k\}_{k=1}^M$ and Δx (Eqn. [31])
4: Initialise Background Lattice: (x_i, t_n)
for all n **do**
  for all j **do**
    5: Set Time and State Steps: $\Delta t_n, \Delta x_n$
    6: Compute Boltzmann Potentials: V_t
    7: Compute Jump Probabilities: r, F_n
    8: Update Sources: $c^{(j,k)}$
    9: Update Boundary Conditions: φ_n^0, φ_n^M
    10: Update Interior Points: φ_n^i (Eqn. [29,20]
    11: Find mid-price: $p_n = \{x_i : \min\{|\varphi_n^i|\}\}$
  end for
end for
  return φ_n^i and p_n on lattice (x_i, t_n)

7.2 Non-uniform fast fourier transform algorithm

Algorithm 2 NUFFT-FFG algorithm

Require:

1.$\boldsymbol{\delta}_i = (\delta_i(I_h))_{h=1}^{n_i-1}$: vector of source strengths for asset i
2. $\tilde{\boldsymbol{t}}^i = (\tilde{t}_h^i)_{h=1}^{n_{i-1}}$ vector of re-scaled sample times for asset $i(\tilde{t}_h^i \in [0, 2\pi])$
3. $M = 2N + 1$: the number of Fourier modes computed
4. ϵ: error tolerance

Step I. Initialisation

I.1. Set: $\sigma = 2$
I.2. Set: $M_r = \sigma M; M_{sp} = \left[\frac{-\ln(\epsilon)(\sigma-1/2)}{(\pi(\sigma-1))} + \frac{1}{2}\right]$
I.3. Set: $\lambda = \frac{\sigma 2 M_{sp}}{\sigma(\sigma-0.5)}; h_x = \frac{2\pi}{M_r}; t_1 = \frac{\pi}{\lambda}$
I.4. Set: $\tau = \frac{\pi\lambda}{M_r^2}$
I.5. Initialise: $(\tilde{f}_l)_{l=1}^{M_r} = \mathbf{0}$, a zero vector of length M_r
for $k = 1$ to M_{sp} **do**
  $E_{3,k} = \exp(-t_1 k^2)$
end for

Step C: Convolution (33)

for $h = 1$ to $n_i - 1$ **do**
  $b_0 = [\frac{\tilde{t}_h^i}{h_x}]$, index of nearest up-sampled grid $\epsilon_{b_0} \leq \tilde{t}_h^i$
  $d = \frac{\tilde{t}_h^i}{h_x} - b_0$
  $E_0 = \mathbf{0}$, a zero vector of length $2M_{sp}$
  $E_1 = e^{-t_1 d^2}; E_{0,M_{sp}} = E_1; E_2 = e^{2t_1 d}$
  for $k = 1$ to M_{sp} **do**
    $E_{0,M_{sp}+k} = E_{3,k} E_1 E_2^k$
  end for
  for $k = 1$ to $M_{sp} - 1$ **do**
    $E_{0,M_{sp}-k} = E_{3,k} E_1 E_2^{-k}$
  end for
  $b_d = \min(M_{sp} - 1, b_0); b_u = \min(M_{sp}, M_r - b_0 - 1)$
  for $k = -M_{sp} + 1$ to $-b_d - 1$ **do**
    $\tilde{f}_{b_0+k+M_r+1} = \tilde{f}_{b_0+k+M_r+1} + \delta_i(I_h) E_{0,M_{sp}+k}$
  end for
  for $k = -b_d$ to b_u **do**
    $\tilde{f}_{b_0+k+1} = \tilde{f}_{b_0+k+1} + \delta_i(I_h) E_{0,M_{sp}+k}$
  end for
  for $k = b_u + 1$ to M_{sp} **do**
    $\tilde{f}_{b_0+k-M_r+1} = \tilde{f}_{b_0+k-M_r+1} + \delta_i(I_h) E_{0,M_{sp}+k}$
  end for
end for

Step F: Compute FFT on over-sampled grid (34)

F.1. Find Fourier coefficients $F_G(dp_i)(k)$ via FFT on the grid $\tilde{f}_l$.

Step D: Deconvolution (35)

Compute: $F(dp_i)(k) = \sqrt{\frac{\pi}{\tau}} e^{k^2\tau} F_G(dp_i)(k) \frac{1}{M_r}$

return $(F(dp_i)(k),\ k \in \{-N, ..., N\})$

8 Reproducing Research

All the code used in this dissertation can be found on GitHub. This code is an extension of three prior projects; the first, a project for modelling the order book as an anomalous diffusion by Diana and Gebbie (2023a) from (Diana and Gebbie, 2023b), and second, the code used to estimate the Epps effect using Fast Fourier Transform methods developed by Chang et al. (2020a) from Chang et al. (2020c), and third, the calibration code developed by Jericevich et al. (2021b) and refined by Dicks and Gebbie (2022). The complete code release used in this project is that of Bauer and Gebbie (2024) as mirrored from the GitHub repository (Bauer, 2024).

The language used is Julia. To run a Julia script you will need Julia to be installed on your system. Once you have done this we recommend installing and using VSCode with the Julia extension. Download the repository from GitHub to your local machine. Use VSCode to open the folder where you have stored the repository files. Thereafter you should be able to go to a Julia file (denoted with the suffix .jl) and click the play button as shown in Picture 8. It will install the necessary dependencies if they have not already been and execute the script.

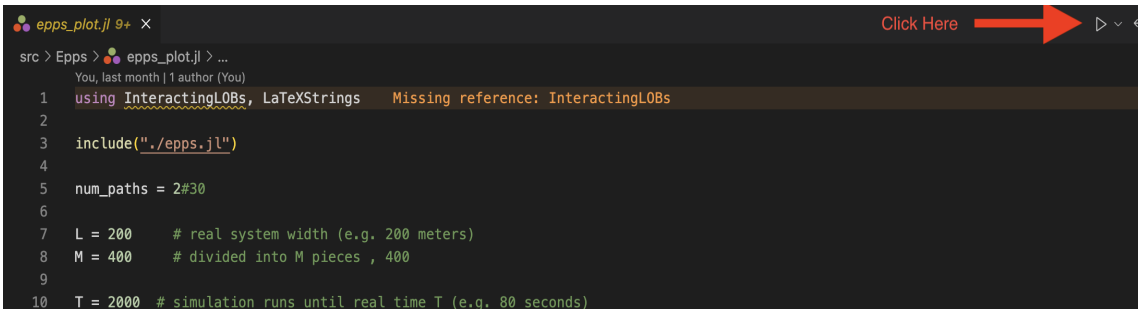


Table 4 shows how each of the figures corresponds to the script as found in the GitHub repository.

Figure	Script file location
Figures 1a & 2	price_surface_with_exp.jl
Figures 1b & 3	price_surface.jl
Figures 4a & 4b	epps_null_case.jl & epps_empirical.jl
Figure 5	epps_initial_plot.jl
Figures 6 & 7	epps_run_all.jl
Figures 8a & 8b	investigate_changing_nu.jl & investigate_changing_nu_non_uniform.jl
Figures 9a & 9b	calibrated_model_epps_non_uniform.jl & calibrated_model_epps_uniform.jl
Figure 10	plot_stylized_facts_calibrated.jl & plot_stylized_facts_empirical.jl

Table 4: A table linking the figures found in the dissertation linked to the script that is needed to reproduce them using the code found at https://github.com/DominicGBauer/InteractingLOBs.jl

References

Angstmann, C., Donnelly, I., Henry, B., Jacobs, B., Langlands, T., Nichols, J., 2016. From stochastic processes to numerical methods: A new scheme for solving reaction subdiffusion fractional partial differential equations. Journal of Computational Physics 307, 508–534. URL: https://www.sciencedirect.com/science/article/pii/S0021999115007937, doi:https://doi.org/10.1016/j.jcp.2015.11.053.

Angstmann, C., Henry, I., Jacobs, B., McGann, A., 2016. Numeric solution of advection-diffusion equations by a discrete time random walk scheme. arXiv: URL: http://arxiv.org/abs/1610.05417v1.

Bauer, D., 2024. Correlation emergence in two coupled limit order books in the fluid limit. MSc. Dissertation. University of Cape Town.

Bauer, D., Gebbie, T., 2024. Correlation emergence in two coupled limit order books in the fluid limit - Julia Code URL: https://zivahub.uct.ac.za/articles/software/Correlation_emergence_in_two_coupled_limit_order_books_in_the_fluid_limit_-_Julia_Code/25151174, doi:10.25375/uct.25151174.v1.

Benzaquen, M., Bouchaud, J.P., 2018. A fractional reaction–diffusion description of supply and demand. The European Physical Journal B 91. URL: http://dx.doi.org/10.1140/epjb/e2017-80246-9, doi:10.1140/epjb/e2017-80246-9.

Bonanno, G., Lillo, F., Mantegna, R.N., 2000. High-frequency cross-correlation in a set of stocks. arXiv:cond-mat/0009350.

Budish, E., Cramton, P., Shim, J., 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. The Quarterly Journal of Economics 130, 1547–1621.

Chang, P., Pienaar, E., Gebbie, T., 2020a. Malliavin–mancino estimators implemented with nonuniform fast fourier transforms. SIAM Journal on Scientific Computing 42, B1378–B1403. doi:10.1137/20M1325903.

Chang, P., Pienaar, E., Gebbie, T., 2020b. The Epps effect under alternative sampling schemes: Dataset. URL: https://zivahub.uct.ac.za/articles/dataset/The_Epps_effect_under_alternative_sampling_schemes_Dataset/13258811, doi:10.25375/uct.13258811.v1.

Chang, P., Pienaar, E., Gebbie, T., 2020c. The Epps effect under alternative sampling schemes: Julia code URL: https://zivahub.uct.ac.za/articles/software/The_Epps_effect_under_alternative_sampling_schemes_Julia_code/13258931, doi:10.25375/uct.13258931.v1.

Chang, P., Pienaar, E., Gebbie, T., 2021a. The epps effect under alternative sampling schemes. Physica A: Statistical Mechanics and its Applications 583, 126329. URL: https://www.sciencedirect.com/science/article/pii/S0378437121006026, doi:https://doi.org/10.1016/j.physa.2021.126329.

Chang, P., Pienaar, E., Gebbie, T., 2021b. Using the epps effect to detect discrete processes. arXiv:2005.10568.

Diana, D., Gebbie, T., 2023a. Anomalous diffusion and price impact in the fluid-limit of an order book. arXiv:2310.06079.

Diana, D., Gebbie, T., 2023b. Simulating Anomalous Diffusion Models and Calculating Price Impacts of The Lit And Latent Order-Book in Julia; code URL: https://zivahub.uct.ac.za/articles/software/Simulating_Anomalous_Diffusion_Models_and_Calculating_Price_Impacts_of_The_Lit_And_Latent_Order-Book_in_Julia_code/22810559, doi:10.25375/uct.22810559.v1.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431. URL: https://doi.org/10.1080/01621459.1979.10482531, doi:10.1080/01621459.1979.10482531, arXiv:https://doi.org/10.1080/01621459.1979.10482531.

Dicks, M., Gebbie, T., 2022. A simple learning agent interacting with an agent-based market model: Julia code URL: https://zivahub.uct.ac.za/articles/software/A_simple_learning_agent_interacting_with_an_agent-based_market_model_Julia_code/21163723, doi:10.25375/uct.21163723.v1.

Dicks, M., Paskaramoorthy, A., Gebbie, T., 2023. A simple learning agent interacting with an agent-based market model. Physica A: Statistical Mechanics and its Applications , 129363URL: https://www.sciencedirect.com/science/article/pii/S0378437123009184, doi:https://doi.org/10.1016/j.physa.2023.129363.

Donier, J., Bonart, J., Mastromatteo, I., Bouchaud, J.P., 2015. A fully consistent, minimal model for non-linear market impact. arXiv URL: https://arxiv.org/abs/1412.0141, arXiv:1412.0141.

Epps, T., 1979. Comovements in stock prices in the very short run. Journal of the American Statistical Association 74, 291–298. doi:10.2307/2286325.

Fabretti, A., 2013. On the problem of calibrating an agent based model for financial markets. Journal of Economic Interaction and Coordination 8, 277–293. URL: https://ideas.repec.org/a/spr/jeicoo/v8y2013i2p277-293.html, doi:10.1007/s11403-012-0096-3.

Gant, M., 2022a. Calibrating a Latent Order Book Model to Market Data. Master's thesis. University of Cape Town, Department of Statistical Sciences. URL: https://open.uct.ac.za/handle/11427/37333.

Gant, M., 2022b. Code: Calibrating a latent order book model to market data. https://github.com/GantZA/SequentialLOB.jl.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1983.tb00371.x, doi:https://doi.org/10.1111/j.1467-9892.1983.tb00371.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9892.1983.tb00371.x.

Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116, 770–799. URL: https://ascelibrary.org/doi/abs/10.1061/TACEAT.0006518, doi:10.1061/TACEAT.0006518, arXiv:https://ascelibrary.org/doi/pdf/10.1061/TACEAT.0006518.

Jericevich, I., Chang, P., Gebbie, T., 2021a. Simulation and estimation of an agent-based market-model with a matching engine. arXiv:2108.07806.

Jericevich, I., Sing, D., Gebbie, T., 2021b. CoinTossX URL: https://zivahub.uct.ac.za/articles/software/CoinTossX/14069552, doi:10.25375/uct.14069552.v1.

Mandelbrot, B.B., Wallis, J.R., 1969. Robustness of the rescaled range r/s in the measurement of noncyclic long run statistical dependence. Water Resources Research 5, 967–988. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR005i005p00967, doi:https://doi.org/10.1029/WR005i005p00967, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR005i005p00967.

Massey, F.J., 1951. The kolmogorov-smirnov test for goodness of fit. Journal of the American Statistical Association 46, 68–78. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500769, doi:10.1080/01621459.1951.10500769, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1951.10500769.

Mastromatteo, I., Marsili, M., Zoi, P., 2010. Financial correlations at ultra-high frequency: Theoretical models and empirical estimation. The European Physical Journal B - Condensed Matter and Complex Systems 80, 243–253. doi:10.1140/epjb/e2011-10865-y.

Mazumder, S., 2016 URL: https://www.sciencedirect.com/science/article/pii/B9780128498941000160, doi:https://doi.org/10.1016/B978-0-12-849894-1.00016-0.

Menkveld, A.J., 2013. High frequency trading and the new-market makers. Journal of Financial Markets 16, 712–740.

Muthuswamy, J., Sarkar, S., Low, A., Terry, E., 2001. Time variation in the correlation structure of exchange rates: high-frequency analyses. Journal of Futures Markets 21, 127–144. doi:https://doi.org/10.1002/1096-9934(200102)21:2<127::AID-FUT2>3.0.CO;2-B.

Münnix, M.C., Schäfer, R., Guhr, T., 2010. Impact of the tick-size on financial returns and correlations. Physica A: Statistical Mechanics and its Applications 389, 4828–4843. URL: https://www.sciencedirect.com/science/article/pii/S0378437110005662, doi:https://doi.org/10.1016/j.physa.2010.06.037.

Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. The Computer Journal 7, 308–313. URL: https://doi.org/10.1093/comjnl/7.4.308, doi:10.1093/comjnl/7.4.308, arXiv:https://academic.oup.com/comjnl/article-pdf/7/4/308/1013182/7-4-308.pdf.

Nuyts, J., 2010. Inference about the tail of a distribution: Improvementon the hill estimator. International Journal of Mathematics and Mathematical Sciences 2010. doi:10.1155/2010/924013.

O'Hara, M., 2015. Market Microstructure Theory. Wiley.

Platt, D., 2020. A comparison of economic agent-based model calibration methods. Journal of Economic Dynamics and Control 113, 103859. URL: https://www.sciencedirect.com/science/article/pii/S0165188920300294, doi:https://doi.org/10.1016/j.jedc.2020.103859.

Platt, D., Gebbie, T., 2017. The problem of calibrating an agent-based model of high-frequency trading. arXiv:1606.01495.

Precup, O.V., Iori, G., 2007. Cross-correlation measures in the high-frequency domain. The European Journal of Finance 13, 319–331. URL: https://doi.org/10.1080/13518470600813565, doi:10.1080/13518470600813565.

Renò, R., 2001. A closer look at the epps effect. International Journal of Theoretical and Applied Finance 06. doi:10.2139/ssrn.314723.

Tumminello, M., Di Matteo, T., Aste, T., Mantegna, R., 2006. Correlation based networks of equity returns sampled at different time horizons. arXiv.org, Quantitative Finance Papers 55. doi:10.1140/epjb/e2006-00414-4.

Tóth, B., Kertész, J., 2009. The epps effect revisited. Quantitative Finance 9, 793–802. URL: https://doi.org/10.1080/14697680802595668, doi:10.1080/14697680802595668, arXiv:https://doi.org/10.1080/14697680802595668.

Tóth, B., Lempérière, Y., Deremble, C., de Lataillade, J., Kockelkoren, J., Bouchaud, J.P., 2011. Anomalous price impact and the critical nature of liquidity in financial markets. Physical Review X 1. URL: http://dx.doi.org/10.1103/PhysRevX.1.021006, doi:10.1103/physrevx.1.021006.

Tóth, B., Tóth, B., Kertész, J., 2007. Modeling the epps effect of cross correlations in asset prices, in: Kertész, J., Bornholdt, S., Mantegna, R.N. (Eds.), SPIE Proceedings, SPIE. URL: http://dx.doi.org/10.1117/12.727127, doi:10.1117/12.727127.

Winker, P., Gilli, M., Jeleskovic, V., 2007. An objective function for simulation based inference on exchange rate data. Journal of Economic Interaction and Coordination 2, 125–145. URL: https://ideas.repec.org/a/spr/jeicoo/v2y2007i2p125-145.html, doi:10.1007/s11403-007-0020-4.

Zebedee, A.A., Kasch-Haroutounian, M., 2009. A closer look at co-movements among stock returns. Journal of Economics and Business 61, 279–294. URL: https://www.sciencedirect.com/science/article/pii/S0148619508000556, doi:https://doi.org/10.1016/j.jeconbus.2008.11.001.