Volatility and Correlation Analysis of Financial Market Data

Georgi Dinolov

University of California, Santa Cruz
Baskin School of Engineering

gdinolov@soe.ucsc.edu

March 11, 2019

Overview

Data type: High-Frequency Prices

- Longitudinal view

Time-stamp	Volume	Price
2011-08-01 03:02:52.112434	1000	179.6400
2011-08-01 03:02:53.752456	1000	179.6200
2011-08-01 03:02:53.900010	1000	179.6300
2011-08-01 03:02:54.103493	1000	179.6050
2011-08-01 03:02:54.343493	1000	179.6700

- Cross-sectional view: the bid-ask spread

Volume	Price
101	179.6400
203	179.6200
305	179.6100
500	179.6000

Data type: Open, Close, High, Low prices

High-frequency returns and difficulties

- As sampling frequency approaches transaction-by-transaction frequency, **irregular spacing** between transactions and **discreteness** in transaction prices (such as the bid-ask spread) become dominant features of the data [10].
- These confounding effects in estimating volatility are generally termed **microstructure noise**, which is
  - independent of sampling interval
  - dominant over short time (< 5 minutes)
  - not cumulative (prices at later times are not affected)

Past work and model-bashed incoherence

- Attempts have been made to fit discrete time state-space models to high-frequency data and estimate price volatility [4, 2]

$$y_{t+\Delta t} = F_t \theta_t + C_t W_{t,1}$$
$$\theta_{t+\Delta t} = G_t \theta_t + V_t W_{t,1}.$$

- As formulated, such types of models are are not robust with respect to the specification of the sampling interval [7, 2, 12].
- Inference over shrinking sampling interval Δt leads to **incoherence**.
- Most recent work has focused on the **realized variance** [6] estimator:

$$RV_T := \sum_{i=1}^{N/\Delta t} r_{t_i}^2,$$

where $r_{t_i}^2$ is the squared *return* observed over $[t_{i-1}, t_i]$.

Realized variance estimator

In the absence of microstructure noise,

$$RV_t = \sum_{i=1}^{N/\Delta t} r_{t_i}^2 \rightarrow \int_0^T \sigma^2 (X_t, t)\, dt$$

- Corrections for the presence of microstructure noise have been proposed in the form of:
  - sampling sparser grids and averaging [11],
  - combining estimators based on sub-sampled data at different frequencies [1],
  - kernel-based estimation [8, 3].
- Disadvantages: averages out information; no inherent model-based projection forward in time.

A filtering-based approach

modeling microstructure noise

Given the **true** price S_j at sample index j, we model the **measured** price P_j as contaminated by noise due to the bid-ask spread D:

$$P_j = S_j + \nu_j, \qquad\qquad \nu_j \sim U[-D/2, D/2],$$

A first-order Taylor expansion of the **observed log price** $\log(P_j)$ produces

$$log(P_j) \approx log(S_j) + \frac{1}{S_j}\nu_j.$$

We further model the noise term with a Gaussian distribution

$$\zeta_j := \frac{1}{S_j}\nu_j \sim N\left(0, \frac{D}{4Q^2}\right)$$

where Q is an order-magnitude approximation of the true price S_j.

Stochastic evolution of price and volatility

the idealized joint log-price $\log(\hat{S}_t)$ and log-volatility $\log(\hat{\sigma}_{t,1}), \log(\hat{\sigma}_{t,2})$ diffusion process as the system of SDEs:

$$
\begin{aligned}
d\log(\hat{S}_t) &= \hat{\mu}\, dt + \sqrt{\hat{\sigma}_{t,1}\hat{\sigma}_{t,2}}\,\sqrt{dt}\hat{\epsilon}_t + dJ_t,\\
d\log(\hat{\sigma}_{t,1}) &= -\hat{\theta}_1(\log(\hat{\sigma}_{t,1}) - \hat{\alpha})\, dt + \hat{\tau}_1\sqrt{dt}\hat{\epsilon}_{t,1},\\
d\log(\hat{\sigma}_{t,2}) &= -\hat{\theta}_2(\log(\hat{\sigma}_{t,2}) - \hat{\alpha})\, dt + \hat{\tau}_2\sqrt{dt}\hat{\epsilon}_{t,2}.
\end{aligned}
$$

- $\log(\hat{\sigma}_{t,1})$ evolves stochastically with long time scale $1/\hat{\theta}_1$,
- $\log(\hat{\sigma}_{t,2})$ evolves stochastically with short time scale $1/\hat{\theta}_2$,
- dJ_t is a compound Poisson process modeling jumps,
- $\hat{\epsilon}_t, \hat{\epsilon}_{t,1}, \hat{\epsilon}_{t,2}$ are Normally distributed with

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

Discrete interpretation of the formulation

Defining the log of the observed price P_j as $Y_j := \log(P_j)$

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

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

$$\alpha(\Delta) = \hat{\alpha} + \frac{1}{2}\log(\Delta), \quad \mu(\Delta) = \hat{\mu}\Delta,$$

$$\tau_i(\Delta) = \hat{\tau}_i\sqrt{\frac{1 - \exp\left\{-2\hat{\theta}_i\Delta\right\}}{2\hat{\theta}_i}}, \quad \theta_i(\Delta) = \exp\left\{-\hat{\theta}_i\Delta\right\}.$$

Prior formulation

To ensure that the model estimation is coherent across choices of sampling periods Δ, we

1. define the first two moments of each continuous-time parameter,
2. use the Delta Method to calculate the first two moments of each discrete-time parameter,
3. define the prior for each discrete-time parameter.

Computation

- We separate the standard innovation term from volatility factors $\sigma_{j,1}$ and $\sigma_{j,2}$ by considering the transformation

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

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

Computation

- We separate the standard innovation term from volatility factors $\sigma_{j,1}$ and $\sigma_{j,2}$ by considering the transformation

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

- We approximate ϵ_j^* as a mixture of Normals

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

Computation, continued

The joint distribution $(\epsilon_j^*, \epsilon_{j,2})$ conditional on the latent mixture element γ_j becomes

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

where d_j is the sign of ϵ_j.

Computation, continued

The joint distribution $(\epsilon_j^*, \epsilon_{j,2})$ conditional on the latent mixture element γ_j becomes

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

where d_j is the sign of ϵ_j. The joint distribution becomes a linear combination of independent Normal random variables when we replace the nonlinear $exp(\epsilon_j^*)$ term with the first-order approximation

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

Blocked Gibbs sampler with predominantly Metropolis-Hastings and Forward Filtering Backward Sampling steps

1. sample observational parameters (MH)
2. sample latent prices (FFBS)
3. sample volatility parameters (MH)
4. sample latent mixture indicators (Discrete)
5. sample volatility paths (FFBS)
6. sample jump parameters (MH)
7. sample jumps (Conjugate prior)

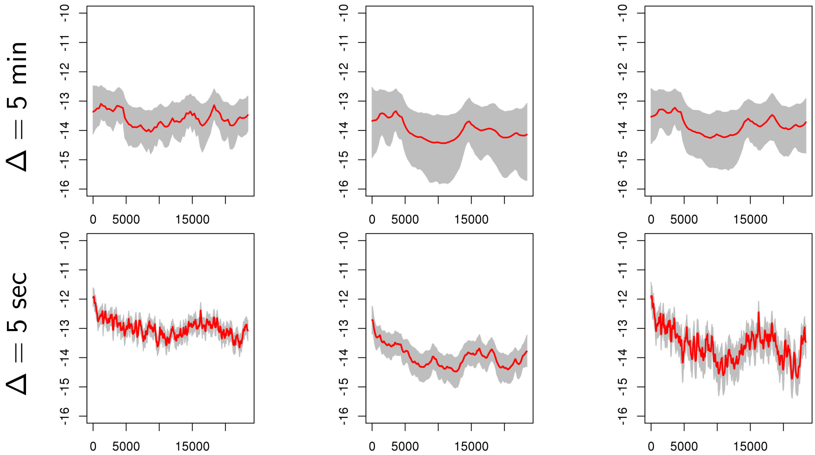
Estimating integrated volatility

- $\zeta \sim N(0, \xi^2)$
- We consider 300 simulated data sets over 1 trading day
- Coverage : percent of 95% probability/confidence intervals covering the true data-generating integrated volatility $\int \hat{\sigma}_t^2 dt$:

	Sampling period			
	60 sec	30 sec	15 sec	5 sec
Inference with $\xi^2 = 0$	72	28	3	0
Inference with ξ^2 fixed at prior mean	79	57	23	0
Inference with ξ^2 estimated	91	92	96	97
Inference with kernel-based estimator	51	48	59	76

Results for real data example

Inference with $\xi^2 = 0$	Inference with ξ^2 fixed at prior mean	Inference with ξ^2 estimated



Log-volatility paths for the AAPL 03/06/2014 data.

Bivariate Open, Close, High, Low Prices

- Work has been done in the univariate case
- No equivalent results in the bivariate setting
- There exists only a single estimator in the literature which used bivariate OCHL to estimate asset correlation [9] (Rogers estimator)

Model formulation

We consider a two-dimensional correlated Brownian motion:

$$X(t) = x_0 + \mu_x t + \sigma_x W_x(t)$$
$$Y(t) = y_0 + \mu_y t + \sigma_y W_y(t)$$

where $W_x(t)$ and $W_y(t)$ are correlated standard Brownian motions with $\mathrm{Cov}(W_x(t), W_x(t)) = \rho t$.

- We seek the 6-dimensional joint probability density function for the pair $(X(t), Y(t))$ and the random variables $M_X(t) = \max_{0\leq s\leq t} X(s)$, $m_X(t) = \min_{0\leq s\leq t} X(s)$, $M_Y(t) = \max_{0\leq s\leq t} Y(s)$, $m_Y(t) = \min_{0\leq s\leq t} Y(s)$.

Transition density

To calculate the joint density

$$p(X(t) = x, Y(t) = y,$$
$$m_X(t) = a_x, M_X(t) = b_X, m_Y(t) = a_Y, M_Y(t) = b_Y),$$

we first study cumulative-like distribution

$$p(X(t) = x, Y(t) = y,$$
$$m_X(t) \geq a_x, M_X(t) \leq b_X, m_Y(t) \geq a_Y, M_Y(t) \leq b_Y),$$

which is governed by a Fokker-Planck equation with absorbing boundaries:

Governing Fokker-Planck equation

Abbreviating

$$q(x,y,t) := p(X(t)=x, Y(t)=y,$$
$$m_X(t) \geq a_x, M_X(t) \leq b_X, m_Y(t) \geq a_Y, M_Y(t) \leq b_Y),$$

the FP equation is

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

$$q(a_x,y,t') = q(b_x,y,t') = q(x,a_y,t') = q(x,b_y,t') = 0, \quad 0 < t' \leq t,$$
$$q(x,y,0) = \delta(x_0 - x)\delta(y_0 - y)$$

Normalized problem

We can introduce a series of transformations to solve the equivalent initial-boundary problem

$$\frac{\partial}{\partial \tilde{t}} q(\tilde{x}, \tilde{y}, \tilde{t}) = \tilde{\mathcal{L}} q(\tilde{x}, \tilde{y}, \tilde{y}), \qquad (\tilde{x}, \tilde{y}) \in \tilde{\Omega}$$

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

where $0 < \sigma_{\tilde{y}}^2 \leq 1$.

- This new parameterization makes the computational domain invariant to data/parameter combinations.

Joint density is given by the fourth derivative with respect to boundaries

Denoting the joint density of the diffusion process and the attained extrema over the interval t as $f(x, y, a_x, b_x, a_y, b_y)$,

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

- The finite difference approximation of $\partial^4 / \partial a_x \partial b_x \partial a_y \partial b_y$ introduces a fundamental limitation due to finite precision round-off errors when the analytic solution $q(x, y, t)$ is not available

$$\mathcal{O}\left(\frac{\varepsilon_{mach}}{\varepsilon^4}\right) \to \infty \quad \text{as} \quad \varepsilon \to 0,$$

where ε is the step size of the numerical differentiation with respect to the boundaries.

Approach 1: Finite Difference approximation to FP equation

Requires the solution to a system of ODEs:

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

where the sparse system matrix B can be computed once and stored for a regular grid over $\tilde{\Omega}$ with finite step size h.

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

Approach 1: Finite Difference approximation to FP equation

Letting $k := 1/h$ and b denote the boundary parameters,

- FD approximation is fast but also limited by irregular truncation errors

$$q^{(k)}(\tilde{x}, \tilde{y}, \tilde{t}|b) - q(\tilde{x}, \tilde{y}, \tilde{t}|b) = \underbrace{\left(\frac{1}{k}\right)^{\alpha} \mathrm{F}_{reg}(b)}_{\text{smooth in } b,\ \alpha>0}$$

$$+ \underbrace{\left(\frac{1}{k}\right)^{\beta} \mathrm{F}_{irreg}(b)}_{\text{continuous but not differentiable w.r.t } b,\ \beta>0} + \underbrace{\varepsilon_{mach} F_{round}(b)}_{\text{behaves as a R.V}}$$

- With linear interpolation which arises when function arguments are not on grid points with respect to boundary differentiation

$$\mathcal{O}(F_{irreg}(b)) = k^{\beta}/\varepsilon \qquad \Rightarrow \left(\frac{1}{k}\right)^{\beta} \mathrm{F}_{irreg}(b) \to \infty \text{ as } \varepsilon \to 0.$$

Approach 2: Trigonometric expansion of the differential operator

$$q(\tilde{x}, \tilde{y}, \tilde{t}) \approx \sum_{\nu} h_{\nu} \phi_{\nu}(\tilde{x}, \tilde{y}) e^{-\lambda_{\nu} \tilde{t}}.$$

- When $\rho = 0$,

$$q(\tilde{x}, \tilde{y}, \tilde{t}) \approx \sum_{l=1}^{L} \sum_{m=1}^{M} c_{l,m} \sin\left(2\pi\, l\, \tilde{x}\right) \sin\left(2\pi\, m\, \tilde{y}\right)$$

  - the system matrix is diagonal
  - no new error is introduced in the time evolution
  - truncated solution converges fast

Approach 2: Trigonometric expansion of the differential operator

- When $\rho \neq 0$, mixing term in the FP equation produces $\cos(2\pi l\tilde{x})\cos(2\pi k\tilde{y})$ terms in the above expansion such that

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

  - system matrix is dense and convergence is slow,
  - new error is introduced in the time evolution

The Galerkin approach: weak solution to the PDE

We propose a solution $q^{(k)}(\tilde{x}, \tilde{y}, \tilde{t})$ of similar form

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

where the basis functions $\psi_i(\tilde{x}, \tilde{y})$ satisfy the boundary conditions on $\tilde{\Omega}$ and they need not be eigenfunctions but capture some essential features of the solution. We also require that all first- and second-order derivatives of $\psi_i(\tilde{x}, \tilde{y})$ are in $L_2(\tilde{\Omega})$.

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

The Galerkin approach: weak solution to the PDE

The *orthogonality* condition of the Galerkin procedure:

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

which is equivalent to the weak formulation of the heat problem

$$\left\langle \partial_t q^{(k)}(\tilde{x},\tilde{y},\tilde{t}),\psi_i \right\rangle = \left\langle \tilde{\mathcal{L}} q^{(k)}(\tilde{x},\tilde{y},\tilde{t}),\psi_i \right\rangle,$$
$$\left\langle q^{(k)}(\tilde{x},\tilde{y},0),\psi_i \right\rangle = \langle q(\tilde{x},\tilde{y},0),\psi_i\rangle,$$

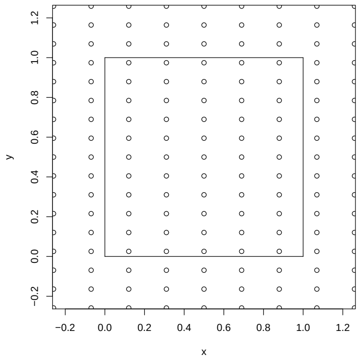
where $< \cdot,\cdot >$ is the usual inner product in $L_2(\tilde{\Omega})$.

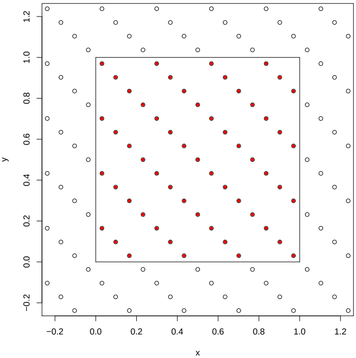
Basis element choice

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

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

- Align basis elements along the principal axis of the differential operator,
- Spacing between nodes is l times the bandwidth $\tilde{\sigma}$ in each principal direction
- This scheme is aimed at better resolving the correlation in the solution

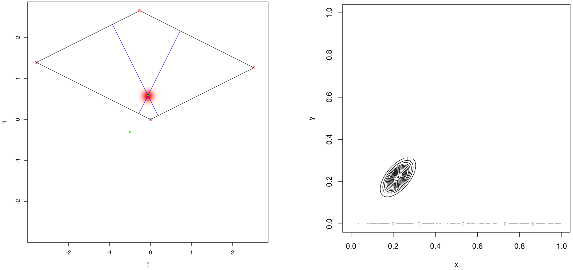




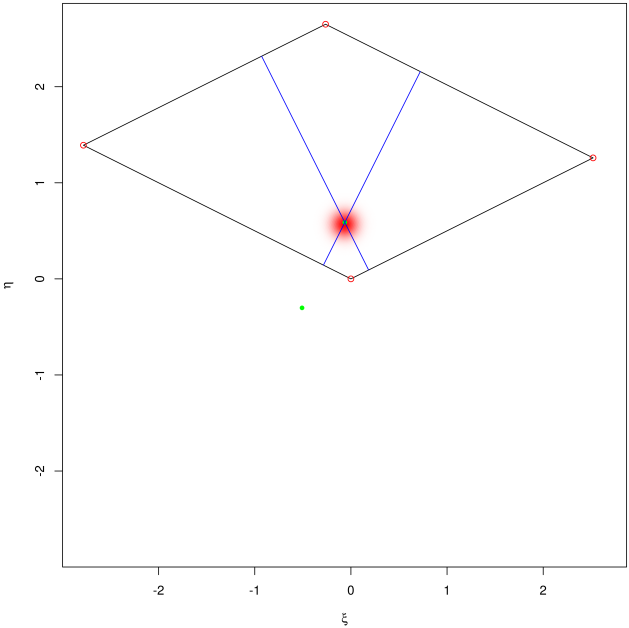
- A sample grid design for $l = 1$, $\tilde{\sigma} = 0.3$ and $\tilde{\rho} = 0.6$.
- **Left:** Grid along coordinates of observed prices within the computational domain (solid black square).
- **Right:** Grid along the major axes of the basis element kernels. The set of final node points $\{(x_i, y_i)\}_{i=0}^{k}$ is contained within the computational domain and is denoted by the red solid points.

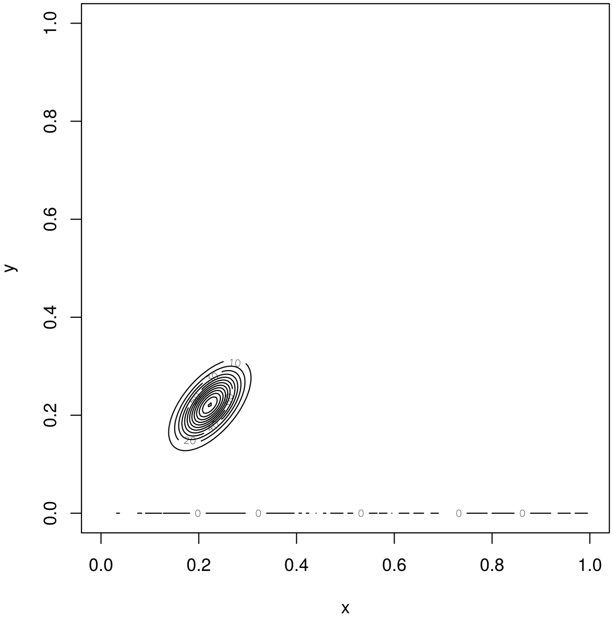
Small-time solution to the Fokker-Planck equation

- A semi-analytic solution for a small time $\tilde{t}_\epsilon$ allows us to project a smooth function onto our basis family instead of a δ-function
- This reduces the numerical error of the Galerkin solver



- **Left:** Computational domain under the transformation where diffusion parameters are both unity and correlation is zero.
- **Right:** Contour of small-time solution in original coordinate system.


2
1
0
-1
-2
η
-2
-1
0
1
2
ξ


y
x
0.0
0.2
0.4
0.6
0.8
1.0
10
20
0

Consistency Results

Denoting the approximate density due to the Galerkin solution as

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

we show that the two integrals below converge in $L_2(\Omega)$:

Lemma (1)

$$\lim_{k \to \infty} \int_{a_x}^{b_x} \int_{a_y}^{b_y} \left(f^{(k)}(x, y, a_x, b_x, a_y, b_y) - f(x, y, a_x, b_x, a_y, b_y) \right)^2 dx\, dy = 0$$

Denoting the random variable

$$Z := (X(t), Y(t), m_X(t), M_X(t), m_Y(t), M_Y(t)),$$

we consider the distribution function $F(Z \leq z)$ generated by the true solution to the Fokker-Planck equation as well as the distribution $F^{(k)}(Z \leq z)$ generated by the Galerkin solution.

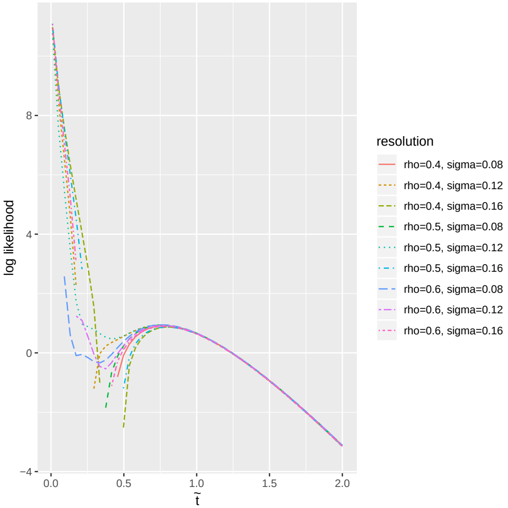
Lemma (Convergence in distribution)

For any $z \in Z$, $\lim_{k\to\infty} F^{(k)}(Z \leq z) = F(Z \leq z)$.

Assuming that the maximum likelihood estimate (MLE) (under $F^{(k)}$ for sufficiently large k) is continuous with respect to the data, we can use the convergence result above and Chebyshev's inequality to show that the MLE under $F^{(k)}$ is consistent as the number of basis elements k and number of data points n go to infinity.

Results and solution behavior for small $\tilde{t}$

Here we consider the parameters defining the basis family $(\tilde{\rho}, \tilde{\sigma})$



Simulation Study

We consider 50 simulated data sets with

$$\sigma_x = 1, \qquad \sigma_y = 1, \qquad \mu_x = 0, \qquad \mu_y = 0$$

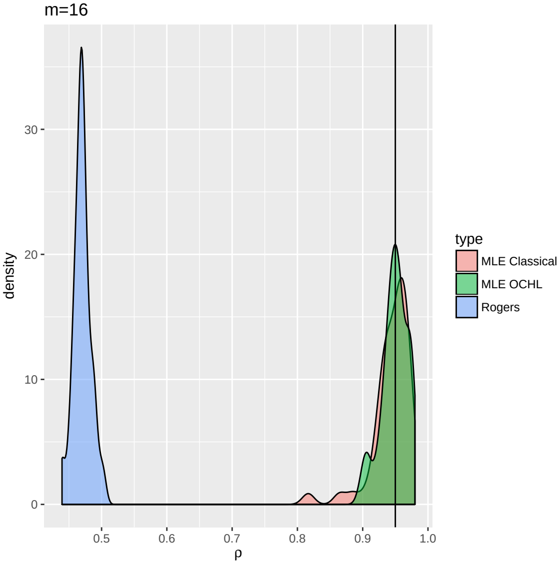
with increasing sample size over several values of ρ. For each case, we consider the ratio of mean-square error of the Galerkin solver compared to that generated by Gaussian likelihoods ignoring the boundaries:

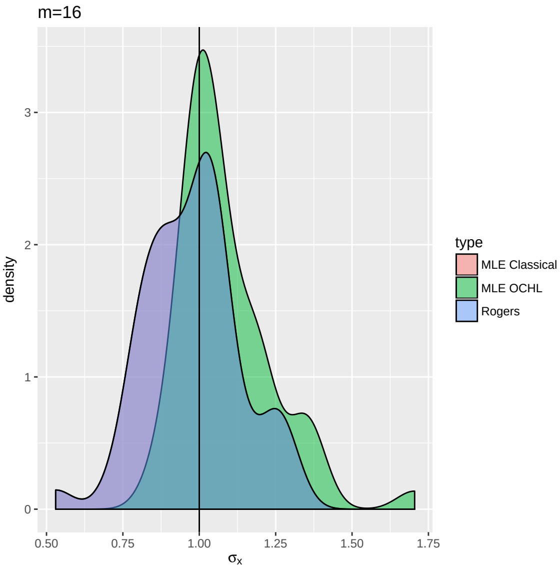
	$\rho = 0.95$			$\rho = 0.60$		
	$m = 4$	$m = 8$	$m = 16$	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.475	1.238	1.304	0.203	0.127	0.232
$\hat{\sigma}_y$	0.593	1.040	1.088	0.111	0.120	0.260
$\hat{\rho}$	0.287	0.910	0.445	0.315	0.283	0.463

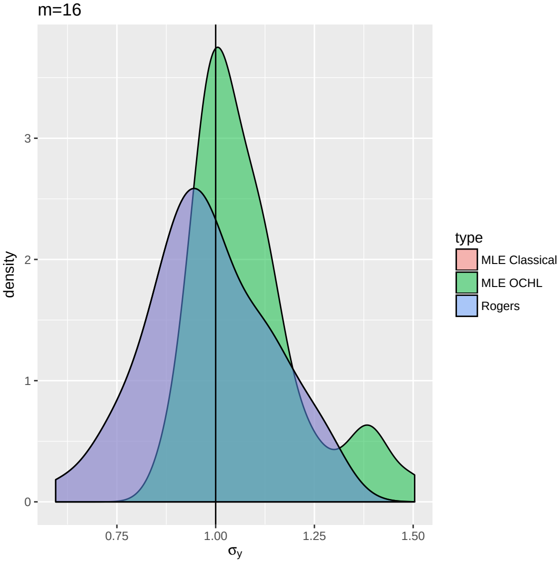
	$\rho = 0.0$		
	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.137	0.243	0.167
$\hat{\sigma}_y$	0.189	0.171	0.107

Simulation study

- For data generated with $\rho = 0.95$, kernel-density approximations of the repeated-sampling densities of the MLEs are shown. Samples are obtained from the Galerkin likelihood (green) and the classical Gaussian likelihood (red) The data-generating parameters are denoted with the vertical solid line.
- The Rogers estimator (blue) is the only existing correlation estimator for bivariate OCHL data.

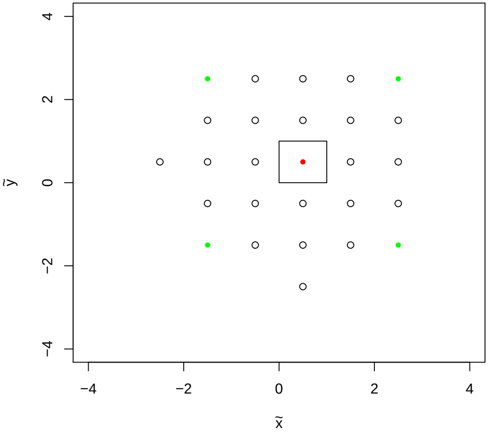

m=16
density
30
20
10
0
0.5
0.6
0.7
0.8
0.9
1.0
ρ
type
MLE Classical
MLE OCHL
Rogers


m=16
density
3
2
1
0
0.50
0.75
1.00
1.25
1.50
1.75
σx
type
MLE Classical
MLE OCHL
Rogers


m=16
density
3
2
1
0
0.75
1.00
1.25
1.50
σy
type
MLE Classical
MLE OCHL
Rogers

Open, Close, High, Low Prices: Small Time Solution and Gap-Fill Solution

Method of images and analytic differentiability



Symmetry constraint in constructing approximate solutions

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

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

Performing the set of reflections

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

produces a sum of images where only four elements are differentiable with respect to all four boundaries:

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

Calculation of the joint density

We can express the derivatives:

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} G(\tilde{x}, \tilde{y}, \tilde{t} | \tilde{x}_{(j)}, \tilde{y}_{(j)}) =$$

$$G \cdot \mathcal{C}^4 \cdot \left(\partial_{a_x} \partial_{b_x} \partial_{a_y} \partial_{b_y} \right) \mathcal{P}$$

$$+ G \cdot \mathcal{C}^3 \cdot \Big(\partial^2_{a_x\ b_x} \partial_{a_y} \partial_{b_y} + \partial^2_{a_x\ a_y} \partial_{b_x} \partial_{b_y} + \partial^2_{a_x\ b_y} \partial_{b_x} \partial_{a_y} + $$

$$+ \partial^2_{b_x\ a_y} \partial_{a_x} \partial_{b_y} + \partial^2_{b_x\ b_y} \partial_{a_x} \partial_{a_y} \partial^2_{a_y\ b_y} \partial_{a_x} \partial_{b_x} \Big) \mathcal{P}$$

$$+ G \cdot \mathcal{C}^2 \cdot \Big(\partial^3_{a_x\ b_x\ a_y} \partial_{b_y} + \partial^2_{a_x\ b_x} \partial^2_{a_y\ b_y} + \partial^3_{a_x\ b_x\ b_y} \partial_{a_y} + $$

$$+ \partial^3_{a_x\ a_y\ b_y} \partial_{b_x} + \partial^2_{a_x\ a_y} \partial^2_{b_y\ b_x} + \partial^3_{b_x\ a_y\ b_y} \partial_{a_x} + \partial^2_{b_x\ a_y} \partial_{a_x} \partial_{b_y} \Big) \mathcal{P}$$

$$+ G \cdot \mathcal{C} \cdot \partial^4_{a_x\ b_x\ a_y\ b_y} \mathcal{P}$$

Analytic Derivative and higher order terms

Thinking of $\tilde{t}$ as variable allows us to further simplify the expression. Since $\mathcal{C} = \mathcal{O}(1/\tilde{t})$, all three terms $G \cdot \mathcal{C}^3$, $G \cdot \mathcal{C}^2$, and $G \cdot \mathcal{C}$ are $o\left(G \cdot \mathcal{C}^4 \cdot \left(\partial_{a_x}\partial_{b_x}\partial_{a_y}\partial_{b_y}\right)\mathcal{P}\right)$, so that the $G \cdot \mathcal{C}^4$ order term in the derivative dominates the others for sufficiently small $\tilde{t}$.

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

Basis Expansion of Galerkin likelihood: RHS

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

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

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

- This suggests a leading-order approximation that is fitted with the Galerkin solver via least squares:

$$f_{\text{Galerkin}}(\tilde{t}) = \tilde{t}^4 \left(\omega_1 e^{-\lambda_1 \tilde{t}} + \omega_2 e^{-\lambda_2 \tilde{t}} \right),$$

$$\log f_{\text{Galerkin}}(\tilde{t}) = \log(\omega_1) + 4 \log(\tilde{t}) - \lambda_1 \tilde{t} + \log \left(1 + \omega_2 / \omega_1 e^{-(\lambda_2 - \lambda_1)\tilde{t}} \right)$$

Leading order for small-time solution: LHS

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

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

- The proposed matched solution is of the form

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

Gap-fill solution

$$\log f_{\text{Galerkin}}(\tilde{t}) \approx \log(\omega_1) + 4\log(\tilde{t}) - \lambda_1\tilde{t} + \log\left(1 + \omega_2/\omega_1 e^{-(\lambda_2-\lambda_1)\tilde{t}}\right)$$

$$\log f_{\text{small-time}}(\tilde{t}) \approx \log(K) - 4.5\log(\tilde{t}) - \frac{\beta_1}{\tilde{t}} + \log(c_1) + \log\left(1 + \epsilon(\tilde{t})\right),$$

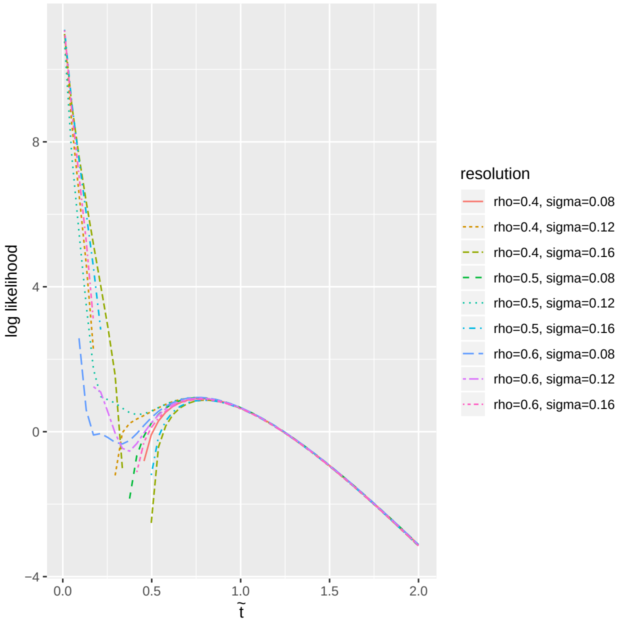
$$\log f_{\text{gap}}(\tilde{t}) = \log(\omega(\tilde{t})) - \gamma(\tilde{t})\log(\tilde{t}) - \frac{\beta(\tilde{t})}{\tilde{t}}$$

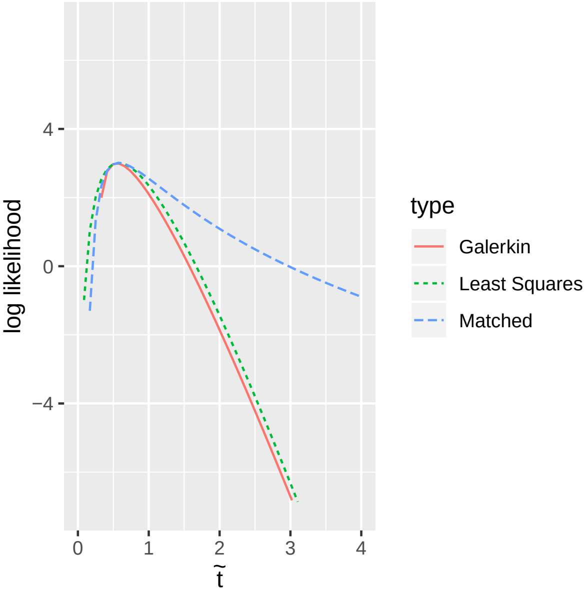
- At $\tilde{t}^*$, the left-hand side of the matching condition, the values for these parameters are defined such that they match the small-time solution

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

- At $\tilde{t}_m$, the right-hand side of the matching condition and the maximum of the LS solution, $\omega(\tilde{t}), \gamma(\tilde{t})$, and $\beta(\tilde{t})$ are chosen to match the value, first, and second derivatives of the logarithmic form of the Galerkin approximate log-likelihood.

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


log likelihood
8
4
0
−4
0.0
0.5
1.0
1.5
2.0
$\tilde{t}$
resolution
rho=0.4, sigma=0.08
rho=0.4, sigma=0.12
rho=0.4, sigma=0.16
rho=0.5, sigma=0.08
rho=0.5, sigma=0.12
rho=0.5, sigma=0.16
rho=0.6, sigma=0.08
rho=0.6, sigma=0.12
rho=0.6, sigma=0.16

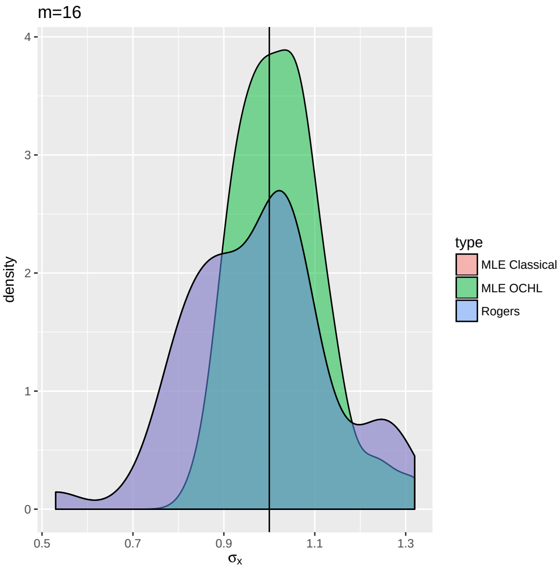


4
0
−4
log likelihood
0
1
2
3
4
$\tilde{t}$
type
Galerkin
Least Squares
Matched

Repeated Simulation Results

$\rho = 0.95$

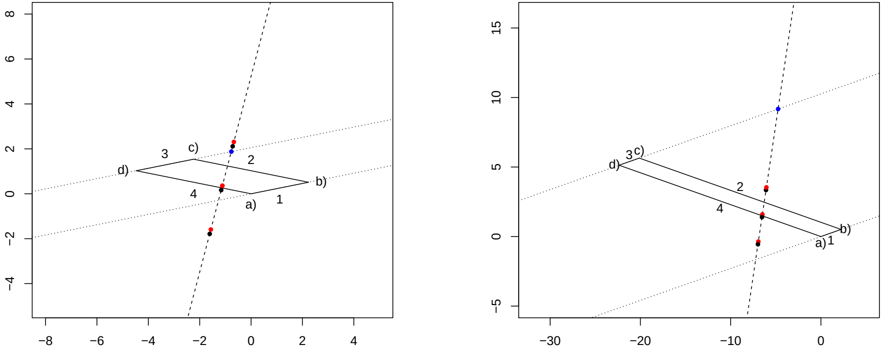
	$m = 4$	$m = 8$	$m = 16$
$\hat{\sigma}_x$	0.124	0.429	0.318
$\hat{\sigma}_y$	0.310	0.147	0.365
$\hat{\rho}$	0.250	0.753	0.699


m=16
density
0
1
2
3
4
0.5
0.7
0.9
1.1
1.3
σx
type
MLE Classical
MLE OCHL
Rogers

- Multivariate High-Frequency models
- Gaussian Process Emulator for interpolating the OCHL likelihood
- Particle Filter

Thank you

Illustration of proof for existence



Prior elicitation example I

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

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

Prior elicitation example II

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

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

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution functions of the standard normal distribution. Now, given the prior mean

Prior elicitation example III

$\hat{a}_{\hat{\theta}}$ and variance $\hat{b}^2_{\hat{\theta}}$ for $\hat{\theta}$, we choose the values of $a_\theta(\Delta)$ and $b_\theta(\Delta)$ so that the mean and variance of $\theta(\Delta)$ above are approximately equal to the mean and variance of $\exp\left\{-\hat{\theta}\Delta\right\}$. To simplify calculation of the moments of $\exp\left\{-\hat{\theta}\Delta\right\}$ we use a second-order Taylor expansion of $\exp\left\{-\hat{\theta}\Delta\right\}$ to approximate the first two moments of $\theta(\Delta)$ in terms of $\hat{a}_{\hat{\theta}}$ and $\hat{b}^2_{\theta}$, an approach known as the Delta-Method (e.g., see [5]):

$$\mathrm{E}\left[\exp\left\{-\hat{\theta}\Delta\right\}\right] \approx \exp\left(-\hat{a}_{\hat{\theta}}\Delta\right)\left(1 + \frac{1}{2}\hat{b}^2_{\hat{\theta}}\Delta^2\right), \tag{3}$$

$$\mathrm{E}\left[\exp\left\{-2\hat{\theta}\Delta\right\}\right] \approx \exp\left(-2\hat{a}_{\hat{\theta}}\Delta\right)\left(1 + 2\hat{b}^2_{\hat{\theta}}\Delta^2\right). \tag{4}$$

Using (1), (2), (3), and (4), and by setting

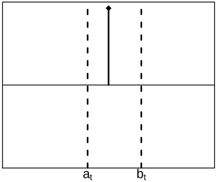
$$\mathrm{E}\left[\theta(\Delta)\right] = \mathrm{E}\left[\exp\left\{-\hat{\theta}\Delta\right\}\right], \qquad \mathrm{Var}\left[\theta(\Delta)\right] = \mathrm{Var}\left[\exp\left\{-\hat{\theta}\Delta\right\}\right],$$

Prior elicitation example IV

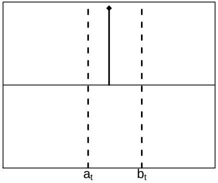
we obtain a system of two equations with two unknowns that can be solved numerically to find the values of $a_\theta(\Delta)$ and $b^2_\theta(\Delta)$ in terms of $\hat{a}_{\hat{\theta}}$, $\hat{b}^2_{\hat{\theta}}$, and Δ.

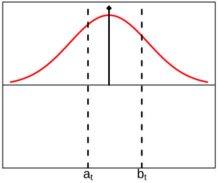
To elicit $\hat{a}_{\hat{\theta}}$ and $\hat{b}^2_{\hat{\theta}}$, recall that $\hat{\theta}$ is the inverse of the time scale of inertia for $\log(\hat{\sigma}_t)$ in the continuous-time formulation, which can be thought of as the characteristic time length, or unit of time, over which the process for the diffusion of $\log(\hat{\sigma}_t)$ “forgets” about an endogenous shock. The two hyper-parameters can be chosen so that the prior probability mass for $\hat{\theta}$ permits a reasonable range for the timescale of inertia.

Solution by Method of Images: single dimension

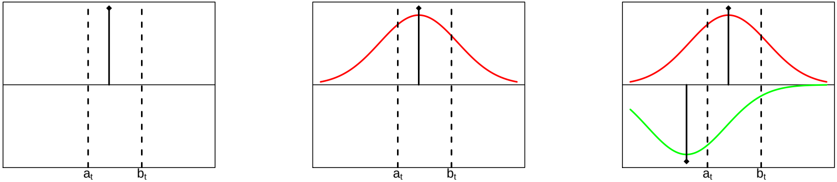


Solution by Method of Images: single dimension

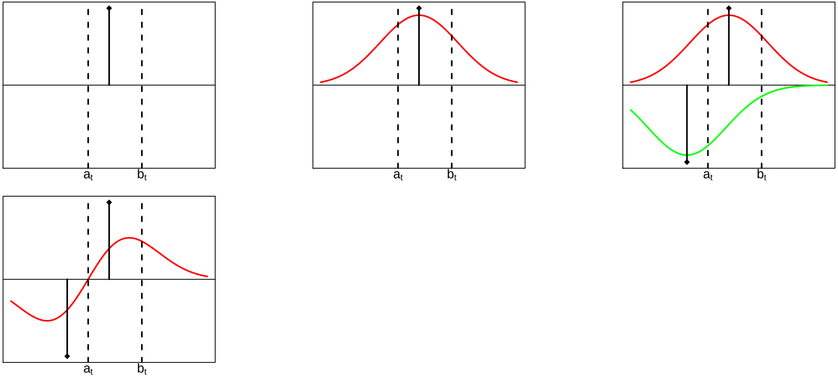




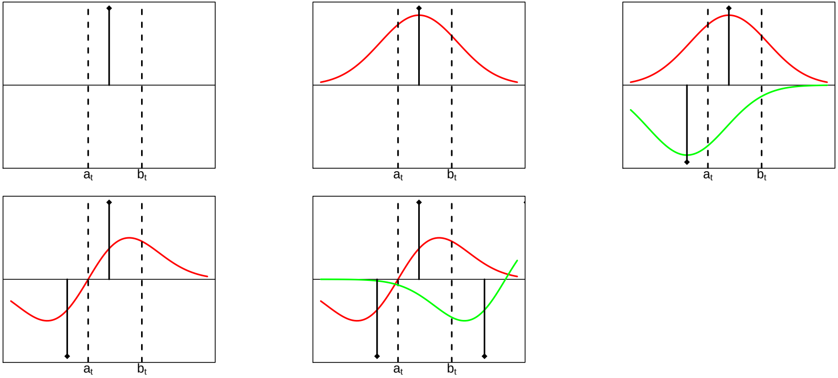
Solution by Method of Images: single dimension



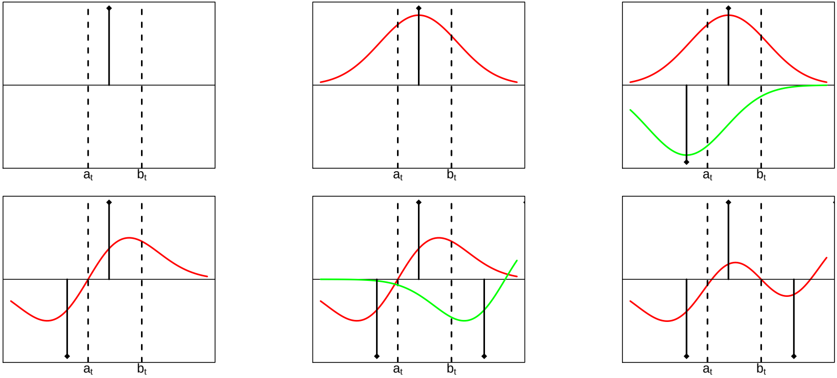
Solution by Method of Images: single dimension



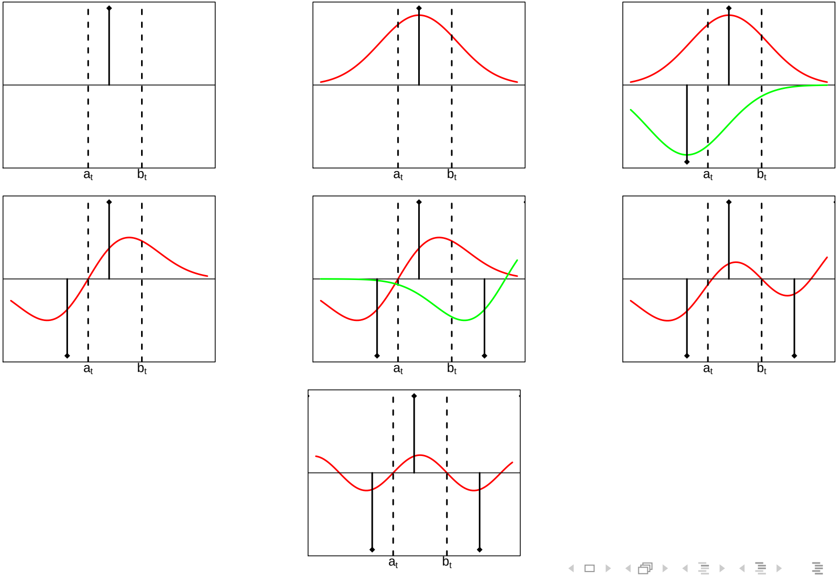
Solution by Method of Images: single dimension



Solution by Method of Images: single dimension

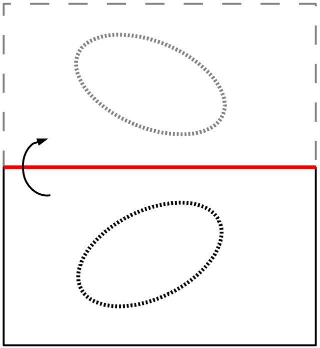


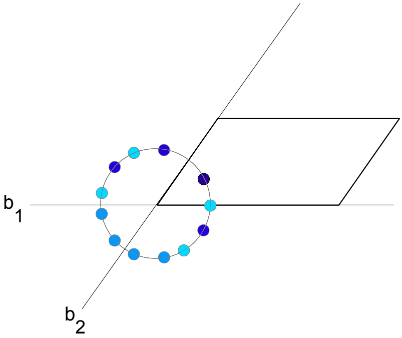
Solution by Method of Images: single dimension



Solution by Method of Images: bivariate case

Method of Images does not work in generate.





References I

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

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

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

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

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

References II

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

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

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

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

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

References III

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

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