Estimation of Volatility Using Open, Close, High, and Low Prices: Application to Stochastic Volatility Models and High-Frequency Data.

Georgi Dinolov

Applied Mathematics and Statistics,
University of California, Santa Cruz

September 11, 2014

Motivation

- Working with returns

$$r_{t+\Delta} \equiv Y_{t+\Delta} - Y_t = \log(S_{t+\Delta}/S_t) \sim N\left(\Delta\mu, \Delta\sigma^2\right).$$

Motivation

- Working with returns

$$r_{t+\Delta} \equiv Y_{t+\Delta} - Y_t = \log(S_{t+\Delta}/S_t) \sim N\left(\Delta\mu, \Delta\sigma^2\right).$$

- Log price $Y_t = \log(S_t)$

$$dY_t = \mu dt + \sigma dW_t$$

Motivation

- Working with returns

$$r_{t+\Delta} \equiv Y_{t+\Delta} - Y_t = \log(S_{t+\Delta}/S_t) \sim N\left(\Delta\mu, \Delta\sigma^2\right).$$

- Log price $Y_t = \log(S_t)$

$$dY_t = \mu dt + \sigma dW_t$$

- Geometric Brownian motion (GBM)

$$dS_t = \mu S_t dt + S_t \sigma dW_t$$

Motivation

- Working with returns

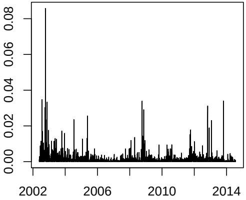
$$r_{t+\Delta} \equiv Y_{t+\Delta} - Y_t = \log(S_{t+\Delta}/S_t) \sim N\left(\Delta\mu, \Delta\sigma^2\right).$$

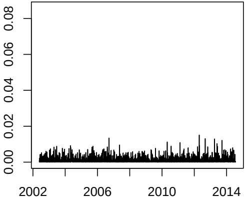
- Log price $Y_t = \log(S_t)$

$$dY_t = \mu dt + \sigma dW_t$$

- Geometric Brownian motion (GBM)

$$dS_t = \mu S_t dt + S_t \sigma dW_t$$





Outline

Outline

Models for Time-Dependent Volatility

1. ARCH/GARCH:

$$\begin{aligned} r_t &= \mu + \epsilon_t & \epsilon_t &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \omega + \alpha\sigma_{t-1}^2 + \beta\epsilon_{t-1}^2 \end{aligned}$$

2. Time-changed Brownian Motion: BM where the time of the process is scaled

$$Y_t = W_{\tau_t},\ \ t \geq 0; \quad Y_t|\tau_t \sim N(0, \tau_t)$$

3. Stochastic volatility:

$$\begin{aligned} Y_t &= \mu + Y_{t-1} + \epsilon_t^1 & \epsilon_t^1 &\sim N(0, \sigma_t^2), \\ \log(\sigma_t) &= \alpha + \phi\left[\log(\sigma_{t-1}) - \alpha\right] + \epsilon_t^2 & \epsilon_t^2 &\sim N(0, \tau^2) \end{aligned}$$

Stochastic Volatility Models

- General case

$$\begin{aligned} dY_t &= \mu(Y_t, \sigma_t)dt + \nu(Y_t, \sigma_t)dW_{t,1} \\ d\sigma_t &= \alpha(Y_t, \sigma_t)dt + \beta(Y_t, \sigma_t)dW_{t,2} \end{aligned}$$

- Continuous time OU

$$\begin{aligned} dY_t &= \mu dt + \sigma_t dW_{t,1} \\ d\log(\sigma_t) &= \underbrace{\phi(\alpha - \log(\sigma_t))dt + \omega dW_{t,2}}_{\text{mean-reverting stochastic process}} \end{aligned}$$

- Discrete time OU

$$\begin{aligned} Y_t &= \mu + Y_{t-1} + \epsilon_t^1 & \epsilon_t^1 &\sim N(0, \sigma_t^2) \\ \log(\sigma_t) &= \alpha + \phi\,[\log(\sigma_{t-1}) - \alpha] + \epsilon_t^2 & \epsilon_t^2 &\sim N(0, \tau^2) \end{aligned}$$

High-Frequency Data and Microstructure Noise

- Availability of tick-by-tick/bid-ask financial transaction data has motivated the development of models that utilize this information to better estimate asset volatility.
- A model-free estimation approach: Realized Volatility

$$RV^{(m)} = \sum_{i=1}^{m} r_i^{(m)^2}$$

It can be shown that $RV^{(m)} \xrightarrow{p} \int_t^{t+\Delta} \sigma^2(Y_s, s)ds$ as $m \to \infty$.

- Problem: microstructure noise on the microsecond level

Outline

Motivation

- GARCH/SV models have been adapted to high-frequency data using RV: subsampling and using RV to summarize intraperiod data.
- We focus on an approach which collapses high-frequency intraperiod transaction information in the form of recorded maximum and minimum prices.

Likelihood for OCHL Data

- We assume constant μ and σ over an interval $s \in [t-1, t]$ with intraperiod min/max a_t and b_t, with BM for the log-price

$$
\begin{aligned}
dY_s &= \mu ds + \sigma dW_s, \\
a_t < \; Y_s &< b_t.
\end{aligned}
$$

Likelihood for OCHL Data

- We assume constant μ and σ over an interval $s \in [t-1, t]$ with intraperiod min/max a_t and b_t, with BM for the log-price

$$\begin{aligned} dY_s &= \mu ds + \sigma dW_s, \\ a_t < \quad Y_s &< b_t. \end{aligned}$$

- Using the Fokker-Planck equation,

$$\begin{aligned} \frac{\partial}{\partial s} q(y,s) &= -\mu \frac{\partial}{\partial y} q(y,s) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial y^2} q(y,s), \\ q(y,t-1) &= \delta(y - y_{t-1}); \quad q(a_t,s) = 0, q(b_t,s) = 0 \end{aligned}$$

$$\boxed{q(y,t) = p(m_t \geq a_t, M_t \leq b_t, Y_t = y | Y_{t-1} = y_{t-1})}$$

Likelihood for OCHL Data

- We assume constant μ and σ over an interval $s \in [t-1, t]$ with intraperiod min/max a_t and b_t, with BM for the log-price

$$\begin{aligned} dY_s &= \mu ds + \sigma dW_s, \\ a_t < \quad Y_s &< b_t. \end{aligned}$$

- Using the Fokker-Planck equation,

$$\begin{aligned} \frac{\partial}{\partial s} q(y,s) &= -\mu \frac{\partial}{\partial y} q(y,s) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial y^2} q(y,s), \\ q(y, t-1) &= \delta(y - y_{t-1}); \quad q(a_t, s) = 0, q(b_t, s) = 0 \end{aligned}$$

$$\boxed{q(y,t) = p(m_t \geq a_t, M_t \leq b_t, Y_t = y | Y_{t-1} = y_{t-1})}$$

- Likelihood for OCHL data

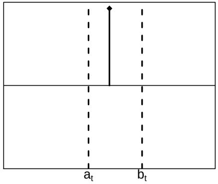
$$\boxed{-\frac{\partial^2}{\partial a_t \partial b_t} q(y,t) = p(m_t = a_t, M_t = b_t, Y_t = y | Y_{t-1} = y_{t-1})}$$

Solution by Fourier Expansion

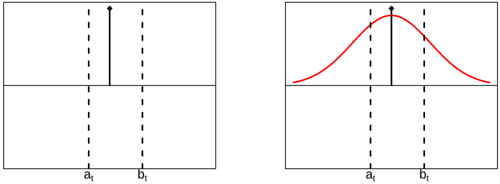
$$q(y,t) = \exp\left(\frac{\sqrt{2\mu}}{\sigma}y\right) \times$$
$$\sum_{n=1}^{\infty} \frac{2}{L} \sin\left((y_{t-1} - a_t)\frac{n\pi}{L}\right) \exp\left(-\frac{1}{2}\sigma^2 \frac{n^2\pi^2}{L^2} t\right) \sin\left((y - a_t)\frac{n\pi}{L}\right)$$

Solution by Method of Images

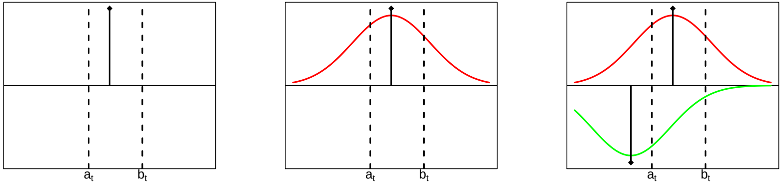
Solution by Method of Images



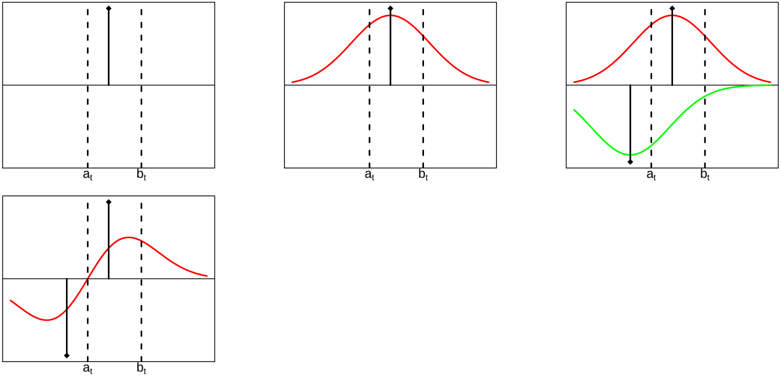
Solution by Method of Images



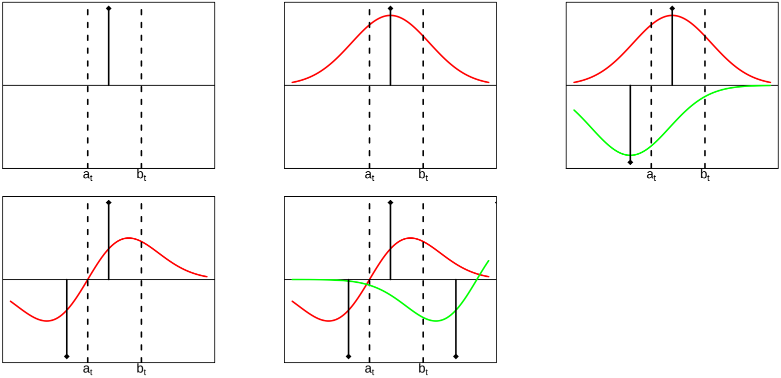
Solution by Method of Images



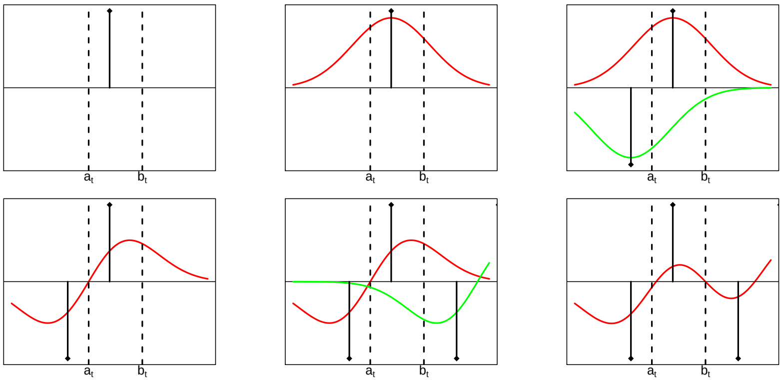
Solution by Method of Images



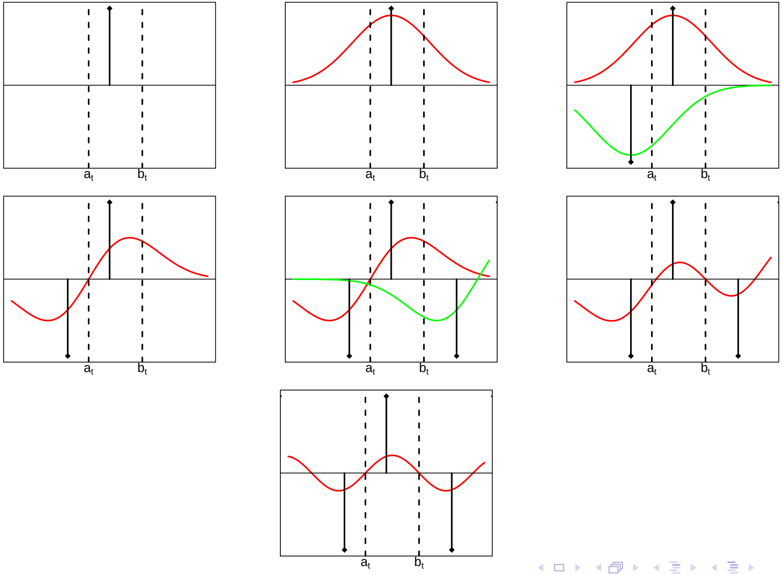
Solution by Method of Images



Solution by Method of Images



Solution by Method of Images



Accuracy of the Fourier Expansion Solution

- Numerical differentiation: central-difference method

$$\Delta x = \frac{1}{2^k}\frac{1}{100}(b_t - a_t).$$

- Analytic differentiation

Accuracy of the Fourier Expansion Solution

- Numerical differentiation: central-difference method

$$\Delta x = \frac{1}{2^k}\frac{1}{100}(b_t - a_t).$$

- Analytic differentiation

Relative error

	N=4	N=8	N=16	N=32	N=64
k=4	-1.290e-3	1.810e-5	1.818e-5	1.818e-5	1.818e-5
k=8	-1.308e-3	7.110e-8	7.111e-8	7.111e-8	7.111e-8
analytic	-1308e-3	-2.002e-14	0	0	0

(N = number of summands in the truncated expansion;
$\mu = 1, \sigma = 1,\ a_t = -0.0557,\ b_t = 3.1424,\ y_t = 3.019,\ y_{t-1} = 0.$)

Simulation

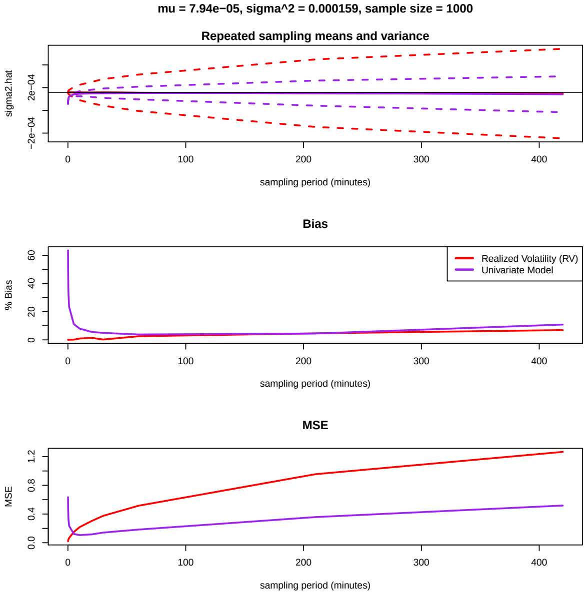
- Goal: compare OCHL estimators to RV estimators
- Forward-Euler discretization

$$Y_{k+1} = Y_k + \mu \Delta t^* + \sigma \Delta t^* \epsilon_k$$

- $\Delta t^* = 1/23400$, equivalent to sampling the returns once every second over the length of a trading day.

$$\mu = 7.936508 \cdot 10^{-5}, \qquad \sigma = 0.01259882$$

- When performing estimation, longer sampling intervals are used: 5 sec to 7 hours.
- No microstructure noise
- 1000 trading days: 1000 estimates for each sampling period Δt used.


mu = 7.94e−05, sigma^2 = 0.000159, sample size = 1000
Repeated sampling means and variance
sigma2.hat
2e−04
−2e−04
0
100
200
300
400
sampling period (minutes)
Bias
% Bias
60
40
20
0
Realized Volatility (RV)
Univariate Model
sampling period (minutes)
MSE
1.2
0.8
0.4
0.0
sampling period (minutes)

Outline

Bivariate Model

- We expect that using OCHL data for two assets can improve estimates for drift and volatility, as well as the estimate of their correlation.

$$\left.\begin{array}{rcl} dY_{1,s} & = & \mu_1 ds + \sigma_1 dW_{1,s}, \\ dY_{2,s} & = & \mu_2 ds + \sigma_2 dW_{2,s} \end{array}\right\} Cor(W_1, W_2) = \rho$$

- Through Fokker-Planck equation, with $\mathbf{y} = (y_1, y_2)$,

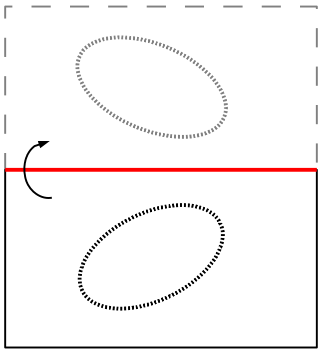
$$\frac{\partial}{\partial s} q(\mathbf{y}, s) = -\mu_1 \frac{\partial}{\partial y_1} q(\mathbf{y}, s) - \mu_1 \frac{\partial}{\partial y_2} q(\mathbf{y}, s)$$

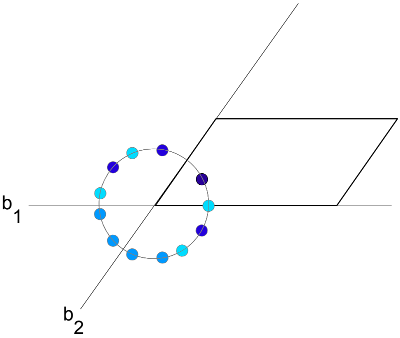
$$+\frac{1}{2}\sigma_1^2 \frac{\partial^2}{\partial y_1^2} q(\mathbf{y}, s) + \frac{1}{2}\sigma_2^2 \frac{\partial^2}{\partial^2 y_2} q(\mathbf{y}, s) + \rho\sigma_1\sigma_2 \frac{\partial^2}{\partial y_1 \partial y_2} q(\mathbf{y}, s)$$

$$\begin{array}{rcl} q(\mathbf{y}, t-1) & = & \delta(y_1 - y_{1,t-1})\delta(y_2 - y_{2,t-1}), \\ q(\mathbf{y}, s) & = & 0 \quad \text{for} \quad y_1 = a_{1,t}, b_{1,t} \quad \text{or} \quad y_2 = a_{2,t}, b_{2,t}. \end{array}$$

Solution through Method of Images

Method of Images does not work.





Solution through Fourier Expansion

$$
\begin{aligned}
q(\mathbf{y}, s) &= \exp(\mathbf{a}^T\mathbf{y} + bs)\times \\
&\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} C_{m,n}(s) \sin\left(\frac{m\pi(y_1 - a_{1,t})}{b_{1,t} - a_{1,t}}\right) \sin\left(\frac{n\pi(y_2 - a_{2,t})}{b_{2,t} - a_{2,t}}\right) \\
&= \exp(\mathbf{a}^T\mathbf{y} + bs)\left(\mathbf{C}(s)^T\mathbf{S}\right)
\end{aligned}
$$

$$
\frac{d}{ds}\begin{pmatrix} C_{1,1}(s) \\ \vdots \\ C_{1,N}(s) \\ C_{2,1}(s) \\ \vdots \\ C_{M,N}(s) \end{pmatrix} = \underbrace{\frac{1}{2}\sigma_1^2\mathbf{A}_1 + \frac{1}{2}\sigma_2^2\mathbf{A}_2 + \rho\sigma_1\sigma_2\mathbf{B}}_{\mathbf{A}} \begin{pmatrix} C_{1,1}(s) \\ \vdots \\ C_{1,N}(s) \\ C_{2,1}(s) \\ \vdots \\ C_{M,N}(s) \end{pmatrix}
$$

Evaluating the Likelihood

$$\frac{\partial^4 q(\mathbf{y}, t)}{\partial a_{1,t} \partial b_{1,t} \partial a_{2,t} \partial b_{2,t}} =$$

$$p(m_{1,t} = a_{1,t}, M_{1,t} = b_{1,t}, m_{2,t} = a_{2,t}, M_{2,t} = b_{2,t}, \mathbf{Y}_t = \mathbf{y} | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})$$

Evaluating the Likelihood

$$\frac{\partial^4 q(\mathbf{y}, t)}{\partial a_{1,t} \partial b_{1,t} \partial a_{2,t} \partial b_{2,t}} =$$

$$p(m_{1,t} = a_{1,t}, M_{1,t} = b_{1,t}, m_{2,t} = a_{2,t}, M_{2,t} = b_{2,t}, \mathbf{Y}_t = \mathbf{y} | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})$$

- Numerical differentiation

$$\frac{\partial^4 q(\mathbf{y}, t)}{\partial a_{1,t} \partial b_{1,t} \partial a_{2,t} \partial b_{2,t}} \approx \frac{\text{finite difference} + \epsilon_{\text{machine}}}{(\Delta x)^4}$$

Evaluating the Likelihood

$$\frac{\partial^4 q(\mathbf{y},t)}{\partial a_{1,t}\partial b_{1,t}\partial a_{2,t}\partial b_{2,t}} =$$

$$p(m_{1,t}=a_{1,t}, M_{1,t}=b_{1,t}, m_{2,t}=a_{2,t}, M_{2,t}=b_{2,t}, \mathbf{Y}_t=\mathbf{y}|\mathbf{Y}_{t-1}=\mathbf{y}_{t-1})$$

- Numerical differentiation

$$\frac{\partial^4 q(\mathbf{y},t)}{\partial a_{1,t}\partial b_{1,t}\partial a_{2,t}\partial b_{2,t}} \approx \frac{\text{finite difference} + \epsilon_{\text{machine}}}{(\Delta x)^4}$$

- Mixed differentiation

$$\frac{\partial^4 q(\mathbf{y},t)}{\partial a_{1,t}\partial b_{1,t}\partial a_{2,t}\partial b_{2,t}} = \exp(\mathbf{a}^T\mathbf{y}+bs)\frac{\partial^4 (\overbrace{\boxed{e^{\mathbf{A}s}\mathbf{C}(0)}}^{\mathbf{C}(s)^T}{}^T\,\mathbf{S})}{\partial a_{1,t}\partial b_{1,t}\partial a_{2,t}\partial b_{2,t}}$$

Numerical Results

- $\Delta x = \min\left\{\frac{1}{2^k}\frac{1}{100}(b_{1,t} - a_{1,t}), \frac{1}{2^k}\frac{1}{100}(b_{2,t} - a_{2,t})\right\}$
- $N = M$ in Fourier expansion
- $\mu_1 = \mu_2 = 1$, $\sigma_1 = \sigma_2 = 1$.

Numerical Results

- $\Delta x = \min\left\{\frac{1}{2^k}\frac{1}{100}(b_{1,t} - a_{1,t}), \frac{1}{2^k}\frac{1}{100}(b_{2,t} - a_{2,t})\right\}$
- $N = M$ in Fourier expansion
- $\mu_1 = \mu_2 = 1$, $\sigma_1 = \sigma_2 = 1$.

Log-likelihood for $\rho = 0.3$

	N=4	N=8	N=16	N=32	N=64
k=3	-1.8300	-1.7475	-1.7601	-1.7416	-1.6371
k=6	-1.8300	-1.7475	-1.7601	-1.7421	-1.6453
k=8	-1.8564	-1.7292	-1.8217	-1.4718	$-\infty$
mixed, k=3	-1.8300	-1.7475	-1.7601	-1.7416	-1.6371
mixed, k=6	-1.8300	-1.7475	-1.7601	-1.7421	-1.6453
mixed, k=8	-1.8300	-1.7475	-1.7601	-1.7416	-1.6365

$(a_{1,t} = -0.7906,\ b_{1,t} = 0.7947,\ y_{1,t} = -0.3946,\ a_{2,t} = -0.02775,\ b_{2,t} = 1.2675,\ y_{2,t} = 1.06421)$

Numerical Results

Log-likelihood for $\rho = 0.5$

	N=4	N=8	N=16	N=32	N=64
k=3	–3.3405	-3.4384	-3.8774	-3.9335	-3.8991
k=6	-3.3409	-3.4354	-3.9105	-3.9596	-3.7720
k=8	-3.3811	-2.9068	-5.4378	$-\infty$	1.0481
mixed, k=3	-3.3405	-3.4384	-3.8776	-3.9338	-3.8981
mixed, k=6	-3.3405	-3.4384	-3.8778	-3.9337	-3.8980
mixed, k=8	-3.3405	-3.4384	-3.8778	-3.9337	-3.8980

$(b_{1,t} = 0.5524,\ a_{1,t} = -0.7481,\ y_{1,t} = -0.3282,\ b_{2,t} = 0.01645,$
$a_{2,t} = -2.7176,\ y_{2,t} = -2.1075\ )$

Numerical Results

Log-likelihood for $\rho = 0$

	N=4	N=8	N=16	N=32	N=64
k=8	-2.1731	-2.1731	-2.1731	-2.1731	-2.1731
mixed, k=8	-2.1731	-2.1731	-2.1731	-2.1731	-2.1731

$(b_{1,t} = 1.5059,\ a_{1,t} = -0.1894,\ y_{1,t} = 1.4004,\ b_{2,t} = 0.5387,\ a_{2,t} = -0.2317,\ y_{2,t} = 0.2569)$

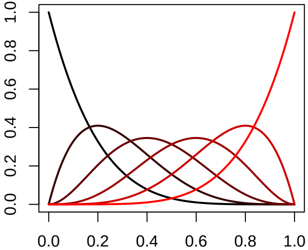
Outline

Bernstein Polynomials

- Sparser matrix

$$B_{\nu,n}(x) = \binom{n}{\nu} x^{\nu}(1-x)^{n-\nu},$$

$$\nu = 0, \ldots, n$$



$$q(\mathbf{y}, s) = \sum_{n=1}^{N} \sum_{m=1}^{M} C_{m,n}(s) B_{n,N}(y_1) B_{m,M}(y_2).$$

- Accounting for correlation

$$q(\mathbf{y}, s) = \sum_{n=1}^{N} \sum_{m=1}^{M} C_{m,n}(s) \mathcal{K}\Big(B_{n,N}(y_1), B_{m,M}(y_2) \Big).$$

- Other bases?

Bivariate Stochastic Volatility Model Using OCHL data

$$\begin{pmatrix} \log(S_{t+1,1}) \\ \log(S_{t+1,2}) \end{pmatrix} = \begin{pmatrix} \log(S_{t,1}) \\ \log(S_{t,2}) \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \mathbf{V_{t+1}} \begin{pmatrix} w_{t+1,1} \\ w_{t+1,2} \end{pmatrix}$$

$$\mathbf{V}_t = \begin{pmatrix} V_{t,11} & 0 \\ V_{t,21} & V_{t,22} \end{pmatrix}, \qquad \mathbf{\Sigma}_t = \mathbf{V}_t' \mathbf{V}_t.$$

$$\begin{pmatrix} \log(V_{t+1,11}) \\ \log(V_{t+1,22}) \\ V_{t+1,21} \end{pmatrix} = \begin{pmatrix} \mu_{11} \\ \mu_{22} \\ \mu_{21} \end{pmatrix} + \mathbf{\Psi} \left(\begin{pmatrix} \log(V_{t,11}) \\ \log(V_{t,22}) \\ V_{t,21} \end{pmatrix} - \begin{pmatrix} \mu_{11} \\ \mu_{22} \\ \mu_{21} \end{pmatrix} \right) + \boldsymbol{\nu},$$

$$\boldsymbol{\nu} \sim N_3(\mathbf{0}, \mathbf{\Sigma}_\nu)$$

Stochastic Volatility Estimation with Microstructure Noise

$$Y_{t+1} = \log(S_{t+1}) + \xi\gamma_t$$

$$\boxed{\begin{aligned} \log(S_{t+1}) &= \log(S_t) + \mu + \sigma_{t+1}\epsilon_{t+1,1} \\ \log(\sigma_{t+1}) &= \alpha + \phi[\log(\sigma_t) - \alpha] + \beta\epsilon_{t+2,2}. \end{aligned}}$$

- $\gamma_t \sim N(0,1)$, $\xi \approx D/2Q$.
- $(\epsilon_{t+1,1}, \epsilon_{t+1,2})$ follows a bivariate normal distribution with $E(\epsilon_{t,1}) = E(\epsilon_{t,2}) = 0$, $\text{Var}(\epsilon_{t,1}) = \text{Var}(\epsilon_{t,2}) = 1$ and $\text{Cor}(\epsilon_{t,1}, \epsilon_{t,2}) = \rho$.

Univariate Likelihood in the Presence of Microstructure Noise

- Our goal is to derive the likelihood for the open, close, high, and low **observed** prices within a period.
- **Observed** intraperiod maximum and minimum are not the **true** maximum and minimum taken on by the asset.
- Previous work is not directly applicable in this setting.
- Brownian motion with jumps?

Timeline for Future Work

	Fall 2014	Winter 2015	Spring 2015	Summer 2015	Fall 2015	Winter 2016
Stochastic Volatility with microstructure noise:						
Univariate Likelihood with microstructure noise:						
Bivariate Likelihood with Bernstein polynomials:						
Bivariate Stochastic Volatility model:						
Thesis writing:						

Questions

Microstructure Variance Specification

- Observe price P_t is the true price S_t plus noise

$$P_t = S_t + \xi^* \epsilon_t, \quad \xi^* = D/2$$

- The log of the observed price is approximated with a first-order Taylor expansion and with average price Q

$$\begin{aligned} \log(P_t) &= \log(S_t + \xi^* \epsilon_t) \\ &\approx \log(S_t) + \frac{\xi^*}{S_t}\epsilon_t \\ &\approx \log(S_t) + \frac{\xi^*}{Q}\epsilon_t \end{aligned}$$

- Hence, the standard deviation in the SV model is approximated with

$$\boxed{\xi = D/2Q}$$

Analytic Matrix Derivatives

- "Matrix" Cauchy Integral

$$f(H) = \frac{1}{2\pi i} \oint f(z)(\mathbf{I}z - \mathbf{H})^{-1} dz$$

- Differentiation with respect to some parameter μ

$$\frac{\partial}{\partial \mu} f(H) = \frac{1}{2\pi i} \oint f(z)(\mathbf{I}z - \mathbf{H})^{-1} \frac{\partial \mathbf{H}}{\partial \mu} (\mathbf{I}z - \mathbf{H})^{-1} dz$$

- Second-order derivative follow

$$\frac{\partial^2}{\partial \mu \partial \nu} f(H) =$$

$$\frac{1}{2\pi i} \oint f(z) \frac{\partial}{\partial \nu} \left[(\mathbf{I}z - \mathbf{H})^{-1} \frac{\partial \mathbf{H}}{\partial \mu} (\mathbf{I}z - \mathbf{H})^{-1} \right] dz$$