

# Simulation of Data

## Simulation of Underlying

### Simulation

These notes discuss how to simulate  $s_{1:t}^*$ . We assume the model

$$dS = rSdt + S\sigma_{LV}(S, t) dB_t \quad (1)$$

and for both simulation and likelihood evaluation we use the Euler approximation

$$s_{t+1} \sim N(s_t + rs_t\Delta_t, \sigma_{LV}^2(s_t, t)s_t^2\Delta_t). \quad (2)$$

This is our recipe for *both*

- simulating  $s_{t+1}|s_t$
- computing likelihoods  $p(s_{t+1}|s_t)$

Spot price paths can be computed independantly of option simulations. Here are some thoughts.

### Implied and Local Volatility

The underlying  $s_t$  requires  $\sigma_{LV}(s_t, t)$  which is the local volatility surface corresponding to the implied volatility surface  $\Sigma(K, T)$ . To compute  $\sigma_{LV}(K, T)$  from  $\Sigma(K, T)$  we can use Dupire's Equation

$$\sigma_{LV}^2(K, T) = \frac{\partial c / \partial T - r(c - K\partial c / \partial K)}{\frac{1}{2}K^2\partial^2 c / \partial K^2} \quad (3)$$

where  $c = c(K, T, \Sigma(K, T))$ . Althternatively we can use the representation in terms of implied volatility  $\Sigma(K, T)$ , which in terms can be evaluated in dollar space or in terms of  $x = \log(s/F)$ . I intend to simulate  $s_{1:t}$  in dollar terms, so I will choose the former. Wilmott (1998) or Kotze Et Al (2015) give the correct expression

$$\sigma_{LV}^2(K, T) = \frac{\Sigma^2 + 2\tau\Sigma\frac{\partial\Sigma}{\partial\tau} + 2rK\tau\Sigma\frac{\partial\Sigma}{\partial K}}{\left(1 + Kd_1\tau\frac{\partial\Sigma}{\partial K}\right)^2 + K^2\tau\Sigma\left(\frac{\partial^2\Sigma}{\partial K^2} - d_1\tau\left(\frac{\partial\Sigma}{\partial K}\right)^2\right)} \quad (4)$$

with the usual definition for  $d_1$ . It may be possible to compute the partials required above but we retain some flexibility using numeric derivatives. If this becomes problematic in terms of efficiency this calculation can be fine tuned. The Dupire approach requires arbitrage free prices, which we can be achieved by either:

- Using a arbitrage-free process such as Fengler (2009)
- Using a surface free of arbitrage by design (eg Gatheral).

though in this project we use the latter.

The Gatheral implied volatility function takes  $x = \log(K/F)$  and defines the implied volatility in terms of the total implied variance  $w(x, t)$  where

$$\begin{aligned} w(x, t) &= \Sigma^2(x, t)t \\ &= \frac{1}{2}\theta_t \left\{ 1 + \rho\varphi(\theta_t)x + \sqrt{(\varphi(\theta_t)x + \rho)^2 + (1 - \rho^2)} \right\} \end{aligned}$$

where  $\theta_t$  is the ATFM total variance, which is determined exogenously from the model either by:

- In live data  $\theta_t$  is computed from option prices

- In simulations  $\theta_t$  is computed from an assumed known model

In simulation experiments I use an initial choice

$$\theta_\tau = \Sigma_0 + \Sigma_1 (1 - \exp(-\lambda_2 \tau)). \quad (5)$$

In theory we should have  $\theta_0 = 0$  but in our experiments we will roll expiries before this limit is reached, allowing us to use  $\Sigma_0 > 0$ .

There are various choices for  $\varphi(\theta_t)$ , but the specific choice used in this project is

$$\varphi(\theta) = \frac{\eta}{\theta^{\gamma_1} (1 + \beta_1 \theta)^{\gamma_2} (1 + \beta_2 \theta)^{1 - \gamma_1 - \gamma_2}}$$

with

$$\begin{aligned} \gamma_1 &= 0.238 \\ \gamma_2 &= 0.253 \\ \beta_1 &= \exp(5.18) \\ \beta_2 &= \exp(-3.00) \\ \eta &= 2.016048 \exp(\varepsilon) \end{aligned}$$

and  $\varepsilon \in (-1, 1)$  and  $\rho$  are parameters. It might make sense to consider the choice  $\gamma_1 = \gamma_2 = 0.25$ , in which case this function will permit more efficient evaluation (eg two `std::sqrt` vs a `std::pow`). In addition the range  $\varepsilon \in (-1, 1)$  my experiments suggest taking a three term approximation to  $\exp(\varepsilon)$  makes no noticeable difference to results.

Gatheral's curve provides us with implied volatilities at each expiry  $\{T_1, T_2, \dots\}$ . However for simulation we require  $\sigma_{LV}(s_t, t)$  where  $t < T_1$  and  $T_i < T_{i+1}$ . Therefore we need to compute the correct value of  $w(x, T_1)$  and extrapolate *backwards* to  $t < T_1$ . But which  $w(x, T_1)$ ? I think  $w(0, T_1)$  is the candidate, but I am open to change. To interpolate I use standard linear interpolation in variance, ie

$$\sigma(t) = \sqrt{\sigma^2(t_0) + \left( \frac{t - t_0}{t_1 - t_0} \right) \sigma^2(t_1)} \quad (6)$$

where  $t_0 < t < t_1$ .