Empirical Asset Pricing through Machine Learning

Hao Xing

Boston University

Introduction

In a Markovian stochastic volatility model, the value function of a European contingent claim is the expectation of the terminal payoff under a (local) martingale measure, conditioning on the market's current configuration. Heuristically, this value function satisfies a PDE, which we call the *valuation equation*. Even though the intuition is clear, it is surprisingly tricky to rigorously prove the heuristic connection, because

- Volatility may vanish on the boundaries of state space. Then the standard Feynman-Kac formula cannot be applied.
- The asset-price process can be a *strict local martingale*; see [1]. Then the valuation equation may have multiple solutions; see [4].

We focus on the following questions in stochastic volatility models:

- What is the concept of a solution (regarding smoothness and boundary conditions) such that the value function is one such solution?
- What is a natural condition under which the value function is the unique solution?
- When the uniqueness in fails, how could one identify the value function among all possible solutions?

The above questions have been discussed in [2], [3] and etc. We assume minimal assumptions on the stochastic volatility models, which may have various behaviors:

- The volatility can potentially reach zero.
- The asset-price can be a strict local martingale.
- The boundary condition to the valuation equation may be unnecessary.
- The payoff function can be unbounded.

The stochastic volatility model

On a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$, the model has the following dynamics:

$$dS_t = S_t\, b(Y_t)\, dW_t, \quad S_0 = x \in \mathbb{R}_+,$$
$$dY_t = \mu(Y_t)\, dt + \sigma(Y_t)\, dB_t, \quad Y_0 = y \in \mathbb{R}_+,$$

where W and B are two Wiener processes with constant correlation $\rho \in (-1, 1)$.

Standing assumption: (satisfied by most models in practice)

(i) $\mu : \mathbb{R}_+ \to \mathbb{R}$ is locally Lip. and $\mu(0) \geq 0$. $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ is locally (1/2)-Hölder, strictly positive on $\mathbb{R}_{++}$, and satisfies $\sigma(0) = 0$. Also,

$$|\mu(y)| + \sigma(y) \leq C(1 + y) \quad \text{for } y \in \mathbb{R}_+.$$

(ii) $b : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies $b(y) > 0$ for $y \in \mathbb{R}_{++}$ and $b(0) = 0$. b is locally α-Hölder, and σb is locally Lip. b has at most polynomial growth.

Zero volatility

For $y \in \mathbb{R}_+$, we define $\tau_0^y := \inf\{t \in \mathbb{R}_{++} \,|\, Y_t^y = 0\}$.

- when $\mu(0) = 0$, $Y_t^y = 0$ for $\tau_0^y \leq t < \infty$, thus 0 is *absorbing*;
- when $\mu(0) > 0$, Y^y is lead back into $\mathbb{R}_{++}$ after τ_0^y.

Lemma: Fix $y \in \mathbb{R}_+$. If $\mu(0) > 0$, then $\int_{\mathbb{R}_+} \mathbb{I}_{\{Y_t^y = 0\}} dt = 0$.

In this case the point 0 is *instantaneously reflecting*.

Martingale property of the asset-price

Let us consider an auxiliary diffusion $\widetilde{Y}$:

$$d\widetilde{Y}_t = (\mu + \rho b \sigma)(\widetilde{Y}_t)\, dt + \sigma(\widetilde{Y}_t)\, dB_t, \quad \widetilde{Y}_0 = y.$$

Proposition: The following statements are equivalent:

(i) $S_{\cdot\wedge T}^{x,y}$ is a strict local martingale for some, then all, $(x, y, T) \in \mathbb{R}_{++}^3$.

(ii) $\widetilde{Y}$ *explodes* to ∞ in finite time.

(iii) $\mathfrak{v}(\infty) = \infty$.

Here $\mathfrak{v}(y) := \int_c^y \frac{\mathfrak{s}(y) - \mathfrak{s}(\xi)}{\mathfrak{s}'(\xi)\sigma^2(\xi)} d\xi$ for $y \in \mathbb{R}_{++}$, where $\mathfrak{s}$ is the scale function for the diffusion $\widetilde{Y}$.

Remark: This proposition uses Sin's argument in [6]. But its statement is stronger than the one in [6]. It says that in a time-homogeneous model, if S is going to lose its martingale property eventually, it must lose it immediately. This proposition also generalizes Theorem 2.4 in [5].

The valuation equation

Given a continuous payoff $g : \mathbb{R}_+^2 \to \mathbb{R}_+$ with $g(x, y) \leq C(1 + x + y^m)$.
The value function of a European option is defined via

$$u(x, y, T) := \mathbb{E}\left[g\left(S_T^{x,y}, Y_T^y\right)\right], \quad \text{for } (x, y, T) \in \mathbb{R}_+^3.$$

u grows at most linear in x and poly. in y.

The valuation equation is

$$\begin{aligned} \partial_T u(x, y, T) &= \mathcal{L}v(x, y, T), \quad (x, y, T) \in \mathbb{R}_{++}^3, \\ u(x, y, 0) &= g(x, y), \quad (x, y) \in \mathbb{R}_+^2, \end{aligned} \tag{VE}$$

in which $\mathcal{L}$ is the infinitesimal generator of (S, Y):

$$\mathcal{L} := \mu(y)\partial_y + \frac{1}{2}b^2(y)x^2\partial_{xx}^2 + \frac{1}{2}\sigma^2(y)\partial_{yy}^2 + \rho b(y)\sigma(y)x\partial_{xy}^2.$$

Note that $\mathcal{L}$ degenerates at $y = 0$ to the operator $\mu(0)\partial_y$.

Boundary conditions:

- When the boundary is absorbing, $v(x, 0, T) = g(x, 0)$ needs to be satisfied.
- When the boundary is instantaneously reflecting, one usually chooses

$$\partial_T v(x, 0, T) = \mu(0)\partial_y v(x, 0, T). \tag{BC}$$

Definition:A function $v \in C(\mathbb{R}_+^3) \cap C^{2,2,1}(\mathbb{R}_{++}^3)$, that solves (VE), is called a *classical solution* in all the following cases (below, y is arbitrary in $\mathbb{R}_{++}$):

(A) When $\mathbb{P}[\tau_0^y = \infty] = 1$.

(B) When $\mathbb{P}[\tau_0^y < \infty] > 0$, $\mu(0) = 0$, and v satisfies $v(x, 0, T) = g(x, 0)$.

(C) When $\mathbb{P}[\tau_0^y < \infty] > 0$, $\mu(0) > 0$, and v belongs to

$$\mathfrak{C} := \left\{ f \in C(\mathbb{R}_+^3) \cap C^{2,2,1}(\mathbb{R}_{++}^3) \,\middle|\, \text{all } \partial_T f,\ \partial_y f,\ y^{2\alpha}\partial_{xx}^2 f \text{ are locally bounded on } \mathbb{R}_{++} \times \mathbb{R}_+ \times \mathbb{R}_{++} \right\}.$$

For any $f \in C(\mathbb{R}_+^3) \cap C^{2,2,1}(\mathbb{R}_{++}^3)$, by saying that $\partial_T f$, $\partial_y f$, and $y^{2\alpha}\partial_{xx}^2 f$ are locally bounded on $y = 0$, we mean that

$$\limsup_{y \downarrow 0} \sup_{(x,T)\in[n^{-1},n]^2} \left[|\partial_T f| + |\partial_y f| + y^{2\alpha}|\partial_{xx}^2 f|\right] < \infty$$

holds for all $n \in \mathbb{N}$.

value-weighted portfolio

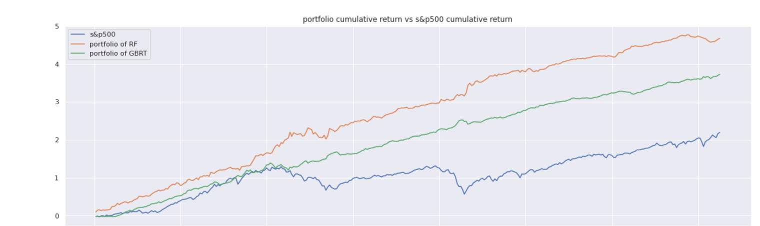
Reasons for value-weighted portfolio construction:

(i) We choose value-weighted portfolios because stocks with a large market value of equity are often important stocks that are traded and held by a large number of people. We want to measure the performance of our models while targeting those important assets.

(ii) The target of our model is to forecast the return of each permno, however, in real life, people often focus on the return of a portfolio rather than single stocks. Thus we also need to measure the performance of simple portfolios to get a better understanding of the performance of our models.

Details of portfolio construction process: We only construct portfolios for gbrt, random forest as well as neural network models as these models have significantly higher out-of-sample R^2 than other models. We set the transaction cost to be 2% and use the S&P500 index as the benchmark. Given the models' predicted returns and the market equity of each permno, we first rank the permnos by their predicted returns, then we select the 50 permnos that have the highest average return as the stocks to buy, and the 50 permnos that have the lowest average return as the stocks to sell, and weight each permno according to their market equity. Doing so, we now have a long-only portfolio and a short-only portfolio for each model, we then combine them to make a long-short portfolio for each model.

Result: while the neural network model has bad performance, the gbrt model and random forest model have relatively good performance and can beat the benchmark.



(a) cumulative return of long-short portfolios

(b) monthly return of long-short portfolios

As we can see, the portfolio based on the predicted returns of random forest model has the hightest cumulative return. However, the sharpe ratio(annualized) of random forest's portfolio is around 1.27, which is smaller then the sharpe ratio of gbrt's portfolio, which is around 3.28.

Table: the SR, IR, MDD of each model

MODELS	annualized sharpe ratio	annualized information ratio (with the benchmark of S&P500 index)	maximum drawdown
neural network	0.599587	0.364212	-3.122637
random forest	3.275392	0.530446	-0.283552
gbrt	1.269521	1.048854	-0.228834

conclusion

Main theorem: The value function $u \in C(\mathbb{R}_+^3) \cap C^{2,2,1}(\mathbb{R}_{++}^3)$. Furthermore, u is the smallest nonnegative classical solution to (VE) in each of the following cases (where y is arbitrary in $\mathbb{R}_{++}$):

(A) When $\mathbb{P}[\tau_0^y = \infty] = 1$.

(B) When $\mathbb{P}[\tau_0^y < \infty] > 0$ and $\mu(0) = 0$.

(C) When $\mathbb{P}[\tau_0^y < \infty] > 0$, $\mu(0) > 0$, and $u \in \mathfrak{C}$.

In all of the above cases, the following two statements hold:

(i) If g is strictly sublinear in x and poly. in y, then u is the unique classical solution within the same class of functions.

(ii) If g is linear growth in x and poly. in y, then u is the unique classical solution within the same class of functions *if and only if* the asset-price process is a martingale.