Jumps and microstructure noise in stock price volatility

Rituparna Sen

Thursday, July 3
Munich
1. Black Scholes model

2 assets:

- Cash Bond

$$\frac{dB_t}{dt} = r_t B_t$$

whose unique solution

$$B_t = B_0 \exp \left( \int_0^t r_s ds \right)$$

$r_t$: short rate, or riskless rate of return, or interest rate, which generally is assumed to be nonrandom, although possibly timevarying.
• Stock

$S_t$: Share price of stock at time $t$

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

(1)

$\{W_t\}_{t \geq 0}$ is a standard Brownian motion,
$\mu_t$ is a nonrandom (but not necessarily constant) function of $t$,
$\sigma > 0$ is a constant called the volatility of the Stock.

Proposition 1. If the drift coefficient function $\mu_t$ is bounded, then the SDE(1) has a unique solution:

$$S_t = S_0 \exp \left( \sigma W_t - \sigma^2 (t/2) + \int_0^t \mu_s ds \right)$$
There exists a measure under which all asset prices are martingales. This is called risk neutral measure.

Under the riskneutral measure, it must be the case that $r_t = \mu_t$.

Corollary 1. Under the riskneutral measure, the log of the discounted stock price $S_T^* = S_t \exp(-\int_0^t r_s ds)$ is normally distributed with mean $\log(S_0) - \sigma^2 t/2$ and variance $\sigma^2 t$.

$$\log(S_T^*) - \log(S_0) \sim \mathcal{N}(\sigma^2 t/2, \sigma^2 t)$$
2. Why volatility?

2.1. Option pricing

A European Call Option on the asset Stock with strike $K$ and expiration date $T$ is a contract that allows the owner to purchase one share of Stock at price $K$ at time $T$. Thus, the value of the Call at time $T$ is $(S_T - K)_+$. According to the Fundamental Theorem of Arbitrage Pricing, the price of the asset Call at time $t = 0$ must be

\[ C(S_0, 0) = C(S_0, 0; K, T) = E(S_T^* - K/B_T)_+ \]
2.2. Portfolio Allocation

Merton’s problem: Want to maximize utility($U$) by investing on the stock and bond. Proportion of wealth to invest in stock is

$$\phi(t) = f(U) \frac{\mu - r}{\sigma^2}$$

where $f$ is a known function of $U$. 
2.3. Risk Management

We want to measure the riskiness of a portfolio. One such measure is value at risk. It is a measure of how much a certain portfolio can lose within a given time period, for a given confidence level. The problem is to estimate a particular quantile of future portfolio values, conditional on current information.

Portfolio returns are functions of $S_t$ whose distribution depends on $\sigma$. 
3. Estimating Volatility

- GARCH Models
- Stochastic Volatility Models
- Range based estimators
- Implied volatility
4. Realized variance

Consider a fixed time period $[0, 1]$ (a trading day, say). Assume availability of $M + 1$ equally spaced observations over the day, $\delta = 1/M$ is the distance between observations. $p_{j\delta}$ is the $j$-th observation on log stock price. $r_{j\delta} = p_{j\delta} - p_{(j-1)\delta}$ is the $j$-th log return.

$$RV = \sum_{1}^{M} r_{j\delta}^2$$

$RV$ is converges to integrated volatility $\int_{0}^{T} \sigma_t^2 dt$ as the sampling frequency goes to $\infty$. (eg. Jacod and Shiryaev)

Andersen et al.(2003) show that $RV$ beats the other estimators mentioned above.
4.1. Microstructure Noise

The \( j \)-th observed logarithmic price is defined as

\[
\tilde{p}_{j\delta} = p_{j\delta} + \eta_{j\delta} \quad j = 0, \ldots, m
\]

Differencing yields

\[
\tilde{p}_{j\delta} - \tilde{p}_{(j-1)\delta} = p_{j\delta} - p_{(j-1)\delta} + \eta_{j\delta} - \eta_{(j-1)\delta}
\]

\[
\tilde{r}_{j\delta} = r_{j\delta} + \epsilon_{j\delta}
\]

Observed return\(=\)True return\(\pm\)Noise
Assumptions:

1. The log price process is a local martingale.
   \[ p_{j\delta} = \int_0^{j\delta} \sigma_t dW_t \]

2. The spot volatility \((\sigma_t)\) is a cadlag process

3. The noise \(\eta\) are i.i.d. mean zero with a bounded 8th moment.

4. The price process and the noise process are independent.
When market microstructure noise is present and unaccounted for, realized variance diverges to infinity. (Bandi and Russell (JFE 2006), Zhang, Mykland, and Ait-Sahalia (JASA 2006)).

Intuition (Let us look at the bias of the estimator)

\[
E \left( \sum \hat{r}_{j\delta}^2 \right) = E \left( \sum r_{j\delta}^2 \right) + E \left( \sum \epsilon_{j\delta}^2 \right) + E \left( \sum \epsilon_{j\delta} r_{j\delta} \right) \\
\approx \int_0^1 \sigma_s^2 ds + M E \left( \eta_{j\delta} - \eta_{(j-1)\delta} \right)^2 \\
= V + 2M \sigma_\eta^2
\]
Figure 1: IBM Signature plot (Andersen et al., 2001)
Most common practice
Use finite sampling frequency like 15 min or 30 min. This involves throwing away a lot of data. Large variance. Cannot rely on theoretical asymptotic convergence results.

Assume parametric structure on volatility and model the noise.

Zhang, Mykland and Ait-Sahalia. forthcoming JASA.
If volatility is a stochastic process with no parametric structure, use subsampling and averaging, involving estimators constructed on 2 time scales.

Barndorff-Neilsen, Hansen, Lunde, Shephard. In preparation. Kernel-based estimators, modified to get rid of end effects, which is necessary for consistency. The optimal estimator
converges to the integrated variance at rate $m^{1/4}$ where $m$ is the number of intraday returns.

- Hansen and Lunde. In Preparation
  Show empirically that microstructure noise is time-dependent and correlated with increments in the efficient price. Apply cointegration techniques to decompose transaction prices and bid-ask quotes into estimate of efficient price and noise.
5. Summary of Volatility Estimation

Volatility is the standard deviation of the log returns (change in value of a financial instrument within a specific time horizon).

Why volatility? Accurate specification of volatility is of crucial importance in several financial and economic decisions such as

- portfolio allocation
- risk management using measures like Value at Risk
- pricing and hedging of derivative securities
Volatility measures

- Squared returns: very noisy.
- Implied volatility: incorporates some price of risk, since it actually measures expected future volatility.
- Realized volatility: consistent estimator of daily integrated volatility if there is no microstructure noise or jumps.
- Bipower variation: consistent for integrated volatility when the underlying price process exhibits occasional jumps.
- Multi-scale estimator: a new realized measure which is consistent for integrated volatility when the prices are contaminated by microstructure noise. However, it has not been studied how this measure performs in the presence of jumps and how to incorporate jumps into their results.
What are functional data?
Univariate Data $\Rightarrow$ Multivariate Data (MDA) $\Rightarrow$ Infinite-dimensional Data (FDA)

Per subject or experimental unit, one samples one or several functions $X(t), t \in \mathcal{T}$ (interval)

Order, neighborhood and smoothness are important concepts in FDA in contrast to Multivariate Data Analysis

Modeling data as independent realizations of a stochastic process with smooth trajectories
7. Decomposing Noisy Functional Data

into a smooth random process $S$ and additive noise:

$$X_{ij} = S_i(t_{ij}) + R_{ij}, \quad i = 1, \ldots n, j = 1, \ldots m.$$ 

$i$ denotes the subject or experimental unit and $j$ denotes the time.

The $R_{ij}$ are additive noise:

- $R_{ij}, R_{i'k}$ are independent for all $i \neq i'$,
- $E(R_{ij}) = 0$, $\text{var}(R_{ij}) = \sigma^2_R < 1$.

Note that the noise $R_{ij}$ within the same subject or item $i$ may be correlated.
8. Modeling Noise Components

\[ \log(R_{ij}^2) = V(t_{ij}) + W_{ij} \]

where \( V \) is the functional variance process which is smooth, i.e., it has a smooth mean function \( \mu_V \), and a smooth covariance structure

\[ G_V(s, t) = \text{cov}(V(s), V(t)), \quad s, t \in \mathcal{T}. \]

The \( W_{ij} \) are 'white noise':

\[ \text{E}(W_{ij}) = 0, \quad \text{var}(W_{ij}) = \sigma_W^2, \]

\[ W_{ij} \perp W_{ik} \quad \text{for} \quad j \neq k, \quad W \perp V, W \perp S. \]
9. Modeling Financial Returns

Black-Scholes model for stock price

\[ d \log X(t, \omega) = \mu dt + \sigma dW(t, \omega), \quad t \geq 0 \]

Here \( W \) is a standard Wiener process, \( \sigma > 0 \) is called the volatility, \( \mu \) is the drift.

Generalizations (Barndorff-Nielsen & Shephard 2002)

\[ d \log X(t, \omega) = \mu(t, \omega) dt + \sigma(t, \omega) dW(t, \omega) \]

where \( \mu(t, \omega), \sigma(t, \omega), W(t, \omega) \) are independent stochastic processes, \( \mu(\cdot), \sigma(\cdot) \) with smooth paths.
10. Volatility Process

For \( X_\Delta(t) = (\log X(t + \Delta) - \log(X(t))) / \sqrt{\Delta} \) and \( W_\Delta(t) = (W(t + \Delta) - (W(t))) / \sqrt{\Delta} \),

obtain in the limit as \( \Delta \to 0 \) \( X_\Delta(t) = \sigma(t)W_\Delta(t) \)

Let \( \eta = \mathbb{E} \log(N(0, 1)^2) \). With \( R(t) = X_\Delta(t) \), obtain

\[
\log(R^2(t)) = \log \sigma^2(t) + \log W^2_\Delta(t) \\
= (\log \sigma^2(t) + \eta) + (\log W^2_\Delta(t) - \eta) \\
= \tilde{V}(t) + \tilde{W}(t)
\]

Let \( V(t) = \tilde{V}(t) - \eta \) providing justification for defining \( V \) as the volatility process.
11. Functional Representation Of Noise Process

The decomposition

\[ Z_{ij} = \log(R_{ij}^2) = V(t_{ij}) + W_{ij} \]

implies

\[ E(Z_{ij}) = E(V(t_{ij})) = \mu_V(t_{ij}) \]

\[ \text{cov}(Z_{ij}, Z_{ik}) = \text{cov}(V_i(t_{ij}), V_i(t_{ik})) = G_V(t_{ij}, t_{ik}), \quad j \neq k \]

for the functional variance process \( V \).
Auto-covariance operator associated with the symmetric kernel $G_V$:

$$G_V(f)(s) = \int_T G_V(s, t)f(t)dt,$$

has smooth eigenfunctions $\psi_k$ with nonnegative eigenvalues $\rho_k$ \Rightarrow\n
Representations

$$G_V(s, t) = \sum_k \rho_k \psi_k(s)\psi_k(t), \quad s, t \in T$$

$$V(t) = \mu_V(t) + \sum_k \zeta_k \psi_k(t),$$

with functional principal component scores $\zeta_k, k \geq 1$ with

$E\zeta_k = 0, \text{var}(\zeta_k) = \rho_k, \zeta_k = \int_T(V(t) - \mu_V(t))\psi_k(t)dt,$

$\zeta_k$ uncorrelated, $\sum \rho_k < \infty$
12. **Functional Principal Component Analysis (FPCA)**

Estimation of mean function $\mu_X$ by cross-sectional averaging, of eigenfunctions and eigenvalues from carefully smoothed covariance surface

$$\Gamma(s, t) = \text{cov}(V(s), V(t))$$

(Rice & Silverman 1991 JRSS-B, Yao et al. 2005 JASA)

**How many components $M$ to include?**

Cross-validation ok, pseudo-AIC, BIC work better.

Asymptotically, $M = M(n) \to \infty$
First step: Smoothing of observed trajectories in the model

\[ X_{ij} = S_i(t_{ij}) + R_{ij} \]

Obtain smoothed trajectories \( \hat{S}_i(t_{ij}) \) and estimated errors and transformed residuals

\[ \hat{R}_{ij} = X_{ij} - \hat{S}_i(t_{ij}) \]

\[ \hat{Z}_{ij} = \log(\hat{R}_{ij}^2) = \log(X_{ij} - \hat{S}_i(t_{ij}))^2 \]

This is finite sample correction and needs to be done because \( \Delta \) is finite.
Second step: Apply functional principal component analysis (Principal Analysis of Random Trajectories - PART algorithm) to the sample of transformed residuals $\hat{Z}_{ij}$:

- Estimate mean function $\mu_V$ (smoothing of cross-sectional averages)
- Estimate smooth covariance surface by smoothing of empirical covariances (omitting the diagonal)
- Obtain eigenvalues/eigenfunctions; choosing number of components $M$ by cross-validation (or AIC, BIC)
- From diagonal of covariance surface, obtain $\text{var}(W_{ij}) = \sigma_W^2$
- Obtain individual FPC scores $\zeta_{ij}$ by integration.
14. Asymptotic Results

Basic Lemma: Under regularity conditions, with smoothing bandwidths $b_S$, $m$ no. of measurements per trajectory,

$$
E \left( \sup_{t \in T} | \hat{S}_i(t) - S_i(t) | \right) = O(b_S^2 + \frac{1}{\sqrt{mb_S}}).
$$

Results for mean and covariance structure:

$$
\sup_{t \in T} | \hat{\mu}_V(t) - \mu_V(t) | = O_p(\alpha_n) \\
\sup_{s,t \in T} | \hat{G}_V(s,t) - G_V(s,t) | = O_p(\beta_n) \\
| \hat{\sigma}_W^2 - \sigma_W^2 | = O_p(\gamma_n)
$$

for sequences $\alpha_n, \beta_n, \gamma_n \to 0$. 
Results for eigenvalues/eigenfunctions of functional variance process:

\[ \sup_{t \in T} | \hat{\psi}_k(t) - \psi_k(t) | \xrightarrow{p} 0 \]
\[ \hat{\rho}_k \xrightarrow{p} \rho_k. \]

Results for individual trajectories of functional variance process:

As \( M = M(n) \to \infty \),

\[ \sup_{1 \leq k \leq M} | \hat{\zeta}_{ik} - \zeta_{ik} | \xrightarrow{p} 0 \]
\[ \sup_{t \in T} | \hat{V}_i(t) - V_i(t) | \xrightarrow{p} 0. \]
15. Jump Detection

For each day \( i \) we calculate \( \Xi_i = \sum_{j=1}^{[T/\Delta]} (\exp\{Y_{ij}\}) \)

Conditioning on the \( V \) process, \( \mathbb{E}(\Xi_i) = \left(1 + \frac{\sigma_W^2}{2}\right) \sum_{j=1}^{[T/\Delta]} (\exp\{V_{ij}\}) \)

\( \text{Var}(\Xi_i) = \exp(V_i)^T G_V \exp(V_i) + \frac{\sigma_W^4}{4} \text{Var}(RV) \)

\( \text{Var}(RV) \) can be estimated using \( \frac{TP}{2[T/\Delta]} \)

\( TP_i = M \mu_{4/3}^{-3} \left( \frac{M}{M-2} \sum_{j=3}^{M} \|r_{i,j-2}\|^{4/3} \|r_{i,j-1}\|^{4/3} \|r_{i,j-1}\|^{4/3} \right) \)

As \( \Delta \) goes to zero, under the null hypothesis of no jumps, the asymptotic distribution of

\[
\frac{\Xi_i - \left(1 + \frac{\sigma_W^2}{2}\right) \sum_{j=1}^{[T/\Delta]} (\exp\{V_{ij}\})}{\sqrt{\exp(V_i)^T G_V \exp(V_i) + \frac{\sigma_W^4 TP_i}{8[T/\Delta]}}} \]

is standard normal.
16. **Real Data Application**

5-minute and 1-minute from November 1997 to March 2006 on

- S&P 500 index
- Japanese Yen and US dollar exchange rate
- Euro and US dollar exchange rate

We have eliminated days when there were trading was thin and the market was open for a shortened session.

Huang and Tauchen study the same instruments over a period from 1997 to 2002. They use five minute data and bipower variation.
Figure 2: Mean function, covariance surface and first three eigenfunctions of S&P 500 data.
Figure 3: Density estimate of $z_{TP}$ (red curve) superimposed with standard normal density (blue curve) for Japanese Yen and US Dollar exchange rate: The left panel uses Huang-Tauchen method. The right panel uses FDA. The figures on top are for 5 min data and those on bottom are for 1 min data.
Figure 4: Density estimate of $z_{TP}$ (red curve) superimposed with standard normal density (blue curve) for Euro and US Dollar exchange rate: The left panel uses Huang-Tauchen method. The right panel uses FDA. The figures on top are for 5 min data and those on bottom are for 1 min data.
17. Components Of Realized Volatility

The realized volatility can be split up into four components:

- the part due to drift,
- smooth time-varying volatility,
- microstructure noise
- jump.

Our approach enables us to separate these four components. Some observations:

- The noise is much higher (high mean) than the smooth part of volatility, but is less variable (low std).
Table 1: Statistics of the four components for S & P 500 1 min data

<table>
<thead>
<tr>
<th>Statistics</th>
<th>log drift</th>
<th>log volatility</th>
<th>log noise</th>
<th>log jump</th>
<th>signed root of nonzero jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>1.0511</td>
<td>0.6500</td>
<td>7.5025</td>
<td>0.3862</td>
<td>2.2817</td>
</tr>
<tr>
<td>median</td>
<td>0.9047</td>
<td>0.5308</td>
<td>7.5121</td>
<td>0.0000</td>
<td>18.0946</td>
</tr>
<tr>
<td>std</td>
<td>0.5693</td>
<td>0.4115</td>
<td>0.1489</td>
<td>1.5619</td>
<td>6.6774</td>
</tr>
<tr>
<td>min</td>
<td>0.1608</td>
<td>0.0760</td>
<td>7.0208</td>
<td>0.0000</td>
<td>-178.8955</td>
</tr>
<tr>
<td>max</td>
<td>3.9068</td>
<td>2.5707</td>
<td>8.0995</td>
<td>10.3736</td>
<td>74.4815</td>
</tr>
<tr>
<td>skewness</td>
<td>1.3273</td>
<td>1.3746</td>
<td>0.0531</td>
<td>3.8660</td>
<td>-1.7736</td>
</tr>
<tr>
<td>kurtosis</td>
<td>5.0608</td>
<td>4.8109</td>
<td>3.2191</td>
<td>16.2645</td>
<td>10.2311</td>
</tr>
</tbody>
</table>

- The drift is high for the index but much lower for exchange rates.
- Jump sizes have a very high variability compared to the other components.
Table 2: Statistics of the four components for JPYA0 1 min data

<table>
<thead>
<tr>
<th>Statistics</th>
<th>log drift</th>
<th>log volatility</th>
<th>log noise</th>
<th>log jump</th>
<th>signed root of nonzero jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0120</td>
<td>0.4742</td>
<td>7.8946</td>
<td>2.4787</td>
<td>4.1936</td>
</tr>
<tr>
<td>median</td>
<td>0.0062</td>
<td>0.4480</td>
<td>7.9257</td>
<td>0.0000</td>
<td>24.2878</td>
</tr>
<tr>
<td>std</td>
<td>0.0154</td>
<td>0.2054</td>
<td>0.2381</td>
<td>3.6183</td>
<td>53.9838</td>
</tr>
<tr>
<td>min</td>
<td>0.0001</td>
<td>0.0965</td>
<td>7.2526</td>
<td>0.0000</td>
<td>-100.8842</td>
</tr>
<tr>
<td>max</td>
<td>0.1083</td>
<td>1.1957</td>
<td>9.1318</td>
<td>9.2636</td>
<td>102.6917</td>
</tr>
<tr>
<td>skewness</td>
<td>2.7286</td>
<td>0.6489</td>
<td>0.9867</td>
<td>0.8055</td>
<td>-0.0985</td>
</tr>
<tr>
<td>kurtosis</td>
<td>12.6244</td>
<td>3.0806</td>
<td>7.5932</td>
<td>1.7226</td>
<td>1.6725</td>
</tr>
</tbody>
</table>
Table 3: Statistics of the four components for EURA0 1 min data

<table>
<thead>
<tr>
<th>Statistics</th>
<th>log drift</th>
<th>log volatility</th>
<th>log noise</th>
<th>log jump</th>
<th>signed root of nonzero jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0173</td>
<td>0.4614</td>
<td>7.7910</td>
<td>1.9009</td>
<td>3.5362</td>
</tr>
<tr>
<td>median</td>
<td>0.0070</td>
<td>0.4475</td>
<td>7.8119</td>
<td>0.0000</td>
<td>25.4217</td>
</tr>
<tr>
<td>std</td>
<td>0.0297</td>
<td>0.1906</td>
<td>0.2028</td>
<td>3.2189</td>
<td>43.6267</td>
</tr>
<tr>
<td>min</td>
<td>0.0004</td>
<td>0.0694</td>
<td>7.2490</td>
<td>0.0000</td>
<td>-95.8760</td>
</tr>
<tr>
<td>max</td>
<td>0.3167</td>
<td>1.0618</td>
<td>8.4426</td>
<td>9.1262</td>
<td>82.6653</td>
</tr>
<tr>
<td>skewness</td>
<td>5.3016</td>
<td>0.4431</td>
<td>0.1896</td>
<td>1.1320</td>
<td>-0.3641</td>
</tr>
<tr>
<td>kurtosis</td>
<td>45.8106</td>
<td>2.7918</td>
<td>3.8489</td>
<td>2.3574</td>
<td>1.8431</td>
</tr>
</tbody>
</table>
18. Conclusions

- We have a way of separating the 4 components and studying them separately. Drift should go away in the limit. Volatility is predictable from previous observations. Microstructure noise is not predictable, but has a more or less fixed level. Jump can be modeled separately as in Tauchen et al as a marked Poisson process.

- Methods that ignore noise essentially put noise and jump components together. However these have entirely different dynamics. Both have low predictability. However, while noise is at a fixed level everyday, jumps are rare and large and might be accompanied by arbitrage opportunities if detected early.
• Anderson et al (2003) estimate jumps happen 3-4 times a year which is consistent with our findings. Since the microstructure noise level goes to infinity, the other methods will ultimately recognize all days as having jumps.

• In practice when information arrives, there is not one single big jump, but a series of small jumps. These have a cumulative effect that is higher than the microstructure noise level, though they might not be very big individually. Our method can capture this kind of behavior.
19. Future Work

- Model the returns as mixed GARCH jump model with AR jump intensity as in Mahue and McCurdy, with 3 components to volatility. The non-parametric estimates obtained in this paper could be used to formulate the prior, or for initial estimates.

- Model the jump process as a marked Poisson process as in Huang and estimate and predict intensity and level of jumps.

- Formulate option pricing in the setting of jump diffusion models and use the non-parametric estimates obtained above to integrate this with implied volatility as in Jiang (2002).

- Apply this to asset allocation, portfolio management as in Liu et al.