Accurate asset price modeling and related statistical problems under microstructure noise

José E. Figueroa-López¹

¹Department of Statistics
Purdue University

Exploring Statistical Science Research Seminar
December 8, 2009
Outline

1 Motivation
   Asset price modeling
   Stochastic models for asset prices

2 High-frequency based statistics
   General idea
   Examples

3 Market Microstructure

4 Open problems
Outline

1 Motivation
   Asset price modeling
      Stochastic models for asset prices

2 High-frequency based statistics
   General idea
   Examples

3 Market Microstructure

4 Open problems
How does the price of a stock behaves in time?
How does the price of a stock behaves in time?
How does the price of a stock behave in time?
How does the price of a stock behave in time?
How does the price of a stock behave in time?
Stylized empirical features of stock prices

1. “Sudden big” changes in the price levels (Jumps)
2. Volatility clustering (intermittency)
3. Log returns with heavy-tails and high-kurtosis distributions
4. Leverage phenomenon
Outline

1 Motivation
   Asset price modeling
   Stochastic models for asset prices

2 High-frequency based statistics
   General idea
   Examples

3 Market Microstructure

4 Open problems
What are good models for the price process $\{S_t\}_{t \geq 0}$?
What are good models for the price process $\{ S_t \}_{t \geq 0}$?

Black-Scholes model (1973); Samuelson (1965):

$$R_i^{(\delta)} = \log \frac{S_{(i+1)\delta}}{S_{i\delta}} = b\delta + \sqrt{\delta} \nu Z_i, \quad \text{where} \quad Z_i \sim \mathcal{N}(0, 1).$$

Jiggling motion, no jump-like changes, no clustering or leverage.
What are good models for the price process $\{S_t\}_{t \geq 0}$?

Stochastic volatility Heston model (1993):

$$R_i^{(\delta)} = b\delta + \sqrt{\delta v_{i-1}^{\delta}} Z_i,$$

$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma \sqrt{\delta v_{i-1}^{\delta}} Z_i', \quad \rho = \text{Corr}(Z, Z').$$
What are good models for the price process \( \{ S_t \}_{t \geq 0} \)?

Stochastic volatility with jumps:

\[
R_i^{(\delta)} = b\delta + \sqrt{\delta \nu_{i-1}^{\delta}} Z_i + \theta \delta^{1/\beta} J_i^{(\beta)},
\]
\[
\nu_i^{\delta} = \nu_{i-1}^{\delta} + \alpha(m - \nu_{i-1}^{\delta})\delta + \gamma \sqrt{\delta \nu_{i-1}^{\delta}} Z_i'.
\]

where \( 0 < \beta < 2 \) and \( J_i^{(\beta)} \) are i.i.d. symmetric with heavy tails:

\[
P(J_i^{(\beta)} \geq x) \sim cx^{-\beta}, \quad \text{as} \quad x \to \infty.
\]

Remarks:

- \( P(Z_i \geq x) \ll e^{-x}/x; \) hence, \( J_i \) feels like jumps compared to \( Z_i; \)
- The larger \( \beta \), the lighter the tails, and the smaller \( J_i^{(\beta)} \) will tend to be;
- \( \beta \) is called the index of jump activity.
What are good models for the price process $\{S_t\}_{t \geq 0}$?

Stochastic volatility with jumps:

$$R_i^{(\delta)} = b \delta + \sqrt{\delta \nu_{i-1}^\delta} \ Z_i + \theta \delta^{1/\beta} J_i^{(\beta)},$$

$$\nu_i^\delta = \nu_{i-1}^\delta + \alpha (m - \nu_{i-1}^\delta) \delta + \gamma \sqrt{\delta \nu_{i-1}^\delta} \ Z'_i.$$
What are good models for the price process \( \{ S_t \}_{t \geq 0} \)?

Stochastic volatility with jumps:

\[
R^{(\delta)}_i = b\delta + \sqrt{\delta \nu^{\delta}_{i-1}} Z_i + \theta \delta^{1/\beta} J^{(\beta)}_i,
\]

\[
\nu^{\delta}_i = \nu^{\delta}_{i-1} + \alpha (m - \nu^{\delta}_{i-1}) \delta + \gamma \sqrt{\delta \nu^{\delta}_{i-1}} Z'_i.
\]
What are good models for the price process $\{S_t\},_{t \geq 0}$?

Stochastic volatility with jumps:

$$R^{(\delta)}_i = b\delta + \sqrt{\delta} \nu_{i-1} \ Z_i + \theta \delta^{1/\beta} J^{(\beta)}_i,$$

$$\nu_i = \nu_{i-1} + \alpha (m - \nu_{i-1}) \delta + \gamma \sqrt{\delta} \nu_{i-1} \ Z'_i.$$
Outline

1 Motivation
   Asset price modeling
   Stochastic models for asset prices

2 High-frequency based statistics
   General idea
   Examples

3 Market Microstructure

4 Open problems
Estimation methods based on high-frequency data

- Set-up:
  - We collect $n$ log returns at intervals of size $\delta$:
    \[
    R_i^{(\delta)} := \log \frac{S_{i\delta}}{S_{(i-1)\delta}},
    \]
    for $i = 1, \ldots, n$.
  - Overnight returns are filtered out;
  - The sampling times are then $t_i = i\delta$ and the time horizon is $T = n\delta$.

- What is it?
  Any estimator $\hat{\theta}^{\delta,n}$ which is consistent for a “parameter” $\theta$ when $\delta \to 0$:
  \[
  \hat{\theta}^{\delta,n} \xrightarrow{P} \theta,
  \]
as $\delta \to 0$, keeping $T$ fixed!
Estimation methods based on high-frequency data

• Set-up:
  • We collect \( n \) log returns at intervals of size \( \delta \):
    \[
    R_i^{(\delta)} := \log \frac{S_{i\delta}}{S_{(i-1)\delta}},
    \]
    for \( i = 1, \ldots, n \).
  • Overnight returns are filtered out;
  • The sampling times are then \( t_i = i\delta \) and the time horizon is \( T = n\delta \).

• What is it?
  Any estimator \( \hat{\theta}_{\delta,n} \) which is consistent for a “parameter” \( \theta \) when \( \delta \to 0 \):
  \[
  \hat{\theta}_{\delta,n} \xrightarrow{P} \theta,
  \]
as \( \delta \to 0 \), keeping \( T \) fixed!
Estimation methods based on high-frequency data

- **Set-up:**
  - We collect \( n \) log returns at intervals of size \( \delta \):
    \[
    R_{i}^{(\delta)} := \log \frac{S_{i\delta}}{S_{(i-1)\delta}},
    \]
    for \( i = 1, \ldots, n \).
  - Overnight returns are filtered out;
  - The sampling times are then \( t_i = i\delta \) and the time horizon is \( T = n\delta \).

- **What is it?**
  Any estimator \( \hat{\theta}^{\delta,n} \) which is consistent for a "parameter" \( \theta \) when \( \delta \to 0 \):
  \[
  \hat{\theta}^{\delta,n} \xrightarrow{P} \theta,
  \]
  as \( \delta \to 0 \), keeping \( T \) fixed!
Outline

1 Motivation
   Asset price modeling
   Stochastic models for asset prices

2 High-frequency based statistics
   General idea
   Examples

3 Market Microstructure

4 Open problems
Examples

1. Quadratic variation of the continuous component:

\[ v_t := \lim_{\delta \to 0} \sum_{i: \delta i \leq t} v_{i-1}^\delta, \quad 0 \leq t \leq T. \]

Then, for any \( \alpha > 0 \) and \( 0 > \omega < 1/2 \),

\[ v_{t, n, \alpha, \omega} := v_t := \sum_{i: \delta i \leq t} |R_i^{(\delta)}|^2 1_{\{|R_i^{(\delta)}| \leq \alpha \delta \omega\}} \xrightarrow{P} v_t, \]

2. Index of jump activity \( \beta \):

\[ \beta_{\delta, n, \alpha, \alpha', \omega} := \beta^\delta := \frac{\log \left( \frac{\sum_{i: \delta i \leq t} |R_i^{(\delta)}|^1 1_{\{|R_i^{(\delta)}| \geq \alpha \delta \omega\}}}{\sum_{i: \delta i \leq t} |R_i^{(\delta)}|^1 1_{\{|R_i^{(\delta)}| \geq \alpha' \alpha \delta \omega\}} } \right)}{\log(\alpha')} \xrightarrow{P} \beta. \]
Index of jump activity for Heston with stable jumps

Stochastic volatility with jumps:

\[
R_i^{(\delta)} = b\delta + \sqrt{\delta \nu_{i-1}^\delta} \, Z_i + \theta \delta^{1/\beta} J_i^{(\beta)},
\]

\[
\nu_i^\delta = \nu_{i-1}^\delta + \alpha (m - \nu_{i-1}^\delta) \delta + \gamma \sqrt{\delta \nu_{i-1}^\delta} \, Z'_i.
\]
Index of Jump Activity for Heston with Stable Jumps

Model:
\[ \begin{align*}
    dX_t &= \xi_t \, dW_t + \theta 
    
    d\xi_t &= \kappa(\eta - \xi_t^2) \, dt + \gamma \xi_t \, dB_t \\
    Z_t \text{ is a } \beta \text{-symmetric Stable with scale 1} \\
    E(dW_t \, dB_t) &= \zeta \, dt
\end{align*} \]

Parameters:
\[ \begin{align*}
    \beta &= 1.5; \quad \eta^{1/2} = 0.25; \quad \gamma = 0.5; \quad \kappa = 5; \quad \zeta = 0.5; \quad \theta = 1; \\
\end{align*} \]
Index of jump activity for Heston model

Stochastic volatility Heston model:

\[ R_i^{(δ)} = bδ + \sqrt{δ} \nu_i^{δ} Z_i, \]
\[ \nu_i^{δ} = \nu_{i-1}^{δ} + α(m - \nu_{i-1}^{δ})δ + γ\sqrt{δ\nu_{i-1}^{δ}} Z_i'. \]
Index of jump activity for Heston model

Index of Activity using $\delta t=5$-sec log return in $T=1$ years

Model:
- $d X_t = \xi_t \, dW_t$, with
  - $d \xi_t = \chi(\eta - \xi_t) \, dt + \gamma \, \xi_t \, dB_t$
- $Z_t$ is a $\beta$ - symmetric Stable with scale 1
  - $E(dW_t \, dB_t) = \zeta \, dt$

Parameters:
- $\eta^{1/2} = 0.25, \gamma = 0.5, \chi = 5, \zeta = -0.5$

Graph showing the index of jump activity for different values of $\omega$: $\omega = 0.3, 0.4, 0.45, 0.5$.
Jump index for Heston with time-changed NIG jumps

Model:

\[
R_i^{(δ)} = b_δ + \sqrt{δ \nu_{i-1}^δ} Z_i + J_i^{(δ\nu_{i-1}^δ)},
\]

\[
\nu_i^δ = \nu_{i-1}^δ + \alpha(m - \nu_{i-1}^δ)δ + \gamma\sqrt{δ\nu_{i-1}^δ} Z_i',
\]

\[
\text{Var}(J_i^{(t)}) = t, \quad J_i^{(t)} \text{ i.i.d. } "Heavy tail distribution", \quad β = 1.
\]
Jump index for Heston with time-changed NIG jumps

Index of Activity using $\delta t=5$-sec log return in $T=3$ years

Model:

$X_t = Z(t) + \omega W_x(t) + \beta U(t) + bt,$

$U(t) = \int_0^t R(u) du,$

$Z(t) = \omega W(U(t)) + \theta U(t),$  

$d\tau(t) = \alpha(m-r(t))dt + \sigma_1(t) dB_1(t).$

$U$ is Inverse Gaussian with $EU(t) = \nu$ t. $E(dW_1 dB_1) = 0.$

Parameter:

$\theta = 0.08, \omega = 0.5, \nu = 0.21, b = 0.1430,$

$\alpha = 1.7630, m = 1, \nu = 0.5630, \omega_x = 0.25.$

Parameter $\omega$ in $\beta_{n, \omega, \omega'}$ ($\omega = 0.25, \delta_n = 1/(252*6.5*60*12)$ years)
Jump index for pure time-changed NIG jumps

Model:

\[
R_i^{(\delta)} = b\delta + J_i^{(\delta v_{i-1}^{\delta})},
\]

\[
\nu_i^\delta = \nu_{i-1}^\delta + \alpha(m - \nu_{i-1}^\delta)\delta + \gamma \sqrt{\delta \nu_{i-1}^\delta} Z_i',
\]

\[
\text{Var}(J_i^{(t)}) = t, \quad J_i^{(t)} \overset{\text{i.i.d.}}{\sim} \text{"Heavy tail distribution"}, \quad \beta = 1.
\]
High-frequency based statistics

Examples

Jump index for pure time-changed NIG jumps

Index of Activity using $\delta t=5$-sec log return in $T=3$ years

Model:

$X_t = Z(t) + bt$,  $\tau(t) = \int_0^t r(u) du$

$Z(t) = c_Z \mathcal{W}(U(t)) + \theta_Z U(t)$,

$dr(t) = \sigma(m-r(t)) dt + \sigma v^{1/2}(t) dB_t$

$U$ is Inverse Gaussian with $E(U(t)) = \nu t$. $E(dW_t, dB_t) = 0$.

Parameter:

$\theta_Z = -0.08$, $\sigma_Z = 0.5$, $\nu = 0.21$, $b = 0.1430$,

$\sigma = 1.7630$, $m = 1$, $\nu = 0.5630$.
Index of jump activity for INTC with 5-sec returns
Index of jump activity for INTC with 5-sec returns


Parameter $\omega$ in $\beta_{n,\omega,\omega'}$ ($\omega=.25, \delta_n=1/(252*6.5*60*12)$ years )
Index of jump activity for INTC with 15-sec returns
Index of jump activity for INTC with 15-sec returns
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   
   (i) Nontrading effects
   
   (ii) Clustering noise
       
       e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.

   (iii) Bid/ask bounce effect
       
       Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
       e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   (iii) Bid/ask bounce effect
       Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:

   (i) Nontrading effects

   (ii) Clustering noise

       e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.

   (iii) Bid/ask bounce effect

       Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
       e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   (iii) Bid/ask bounce effect
       Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   
   (i) Nontrading effects
   
   (ii) Clustering noise
   
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.

   (iii) Bid/ask bounce effect
   
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise

\[ R_{i}^{\text{Obs}} = R_{i}^{(\delta)} + \varepsilon_{i}, \quad \text{where} \quad \varepsilon_{i} \quad \text{is a stationary sequence.} \]
Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.

2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
       e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   (iii) Bid/ask bounce effect
       Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.

3. One modeling approach: Microstructure noise

\[ R_{i}^{\text{Obs}} = \text{Quantization}(R_{i}^{(\delta)} + \varepsilon_{i}), \quad \text{where} \quad \varepsilon_{i} \quad \text{is a stationary sequence.} \]
Estimation under microstructure “noise”

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.

2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.

3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).

4. Need to devise methods that are robust against marker microstructure noise.

5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.
Estimation under microstructure “noise”

1. **Ultra high-frequency sampling will eventually recover the tick-by-tick data.**

2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.

3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).

4. Need to devise methods that are robust against marker microstructure noise.

5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.
Estimation under microstructure “noise”

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.

2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.

3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).

4. Need to devise methods that are robust against marker microstructure noise.

5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.
Estimation under microstructure “noise”

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.

2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.

3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).

4. Need to devise methods that are robust against marker microstructure noise.

5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.
Estimation under microstructure “noise”

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.

2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.

3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).

4. Need to devise methods that are robust against marker microstructure noise.

5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.
Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.
Empirical distribution of returns

Return during a given time period = $\log \frac{\text{Final price}}{\text{Initial price}}$.

Figure 1: Empirical density of one-hour returns (Bayer) vs. density of fitted hyperbolic (blue) and fitted normal distribution (red).
Dynamics of the price process

FIGURE 1.2: Evolution of SLM (NYSE), January-March 1993, compared with a scenario simulated from a Black-Scholes model with same annualized return and volatility.
Times series of returns

Figures taken from Cont (2001)

Five-minute log-return for Yen/Deutschemark exchange rate, 1992-1995

BMW daily log-returns